A DISSERTATION
SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN COMPUTER SCIENCE AND ENGINEERING

From Single-channel EEG to Multimodal Fusion: Advancing Sleep Stage Classification with Graph Neural Network

by

LI Menglei

March 2024

The thesis titled

From Single-channel EEG to Multimodal Fusion: Advancing Sleep Stage Classification with Graph Neural Network

by

LI Menglei

is reviewed and approved by:

Chief referee

Professor Date

ZHAO Qiangfu ZHAO Qiangfu Feb, 19, 2024

Professor Date

LIU Yong LIU Yong Feb. 19, 2024

Professor Date

COHEN Michael Michael Cohen 19 Feb 2024

Senior Associate Professor Date

ZHU Xin Zhu Xin Feb. 19, 2024

THE UNIVERSITY OF AIZU

March 2024

Contents

List of Figures

List of Tables

List of Abbreviations

AASM	American Academy of Sleep Medicine
ASA	American Sleep Association
ASGCN	Actional-structural Graph Convolutional Network
AUC	Area Under Curve
BiLSTM	Bi-directional Long Short-Term Memory
CA	Coordinate Attention
CDH	Central Disorders of Hypersomnolence
CNN	Convolutional Neural Network
CSA	Central Sleep Apnea
CT	Computerized Tomography
DE	Differential Entropy
EEG	Electroencephalogram
EMG	Electromyography
EOG	Electrooculogram
FC	Fully-connected
FN	False Negative
FP	False Positive
FPR	False Positive Rate
GCN	Graph Convolutional Network
GNN	Graph Neural Network
GRU	Gated Recurrent Unit
HAR	Human Activity Recognition
HRV	Heart Rate Variability
ICSD	International Classification of Sleep Disorders
IoMT	Internet of Medical Things
LSTM	Long Short-Term Memory
MLP	Multi-layer Perception
MRI	Magnetic Resonance Imaging
MSTGCN	Multi-view Spatial-temporal Graph Convolutional Network
NLP	Natural Language Processing
NREM	Non-rapid eye Movement
OSA	Obstructive Sleep Apnea
PET	Positron Emission Tomography
PFST	Portuguese Foundation for Science and Technology
PLMD	Periodic Limb Movement Disorder
PSG	Polysomnography
RAM	Recurrent Attention Model
RBD	Rapid Eye Movement Sleep Behavior Disorder

REM	Rapid Eye Movement
RF	Random Forest
R&K	Rechtschaffen and Kales
ROC	Receiver Operating Characteristic
RLS	Restless Legs Syndrome
RNN	Recurrent Neural Network
SENet	Squeeze-and-excitation Network
SRMD	Sleep-related Movement Disorder
ST-GCN	Spatiotemporal Graph Convolutional Network
STN	Spatial Transformer Network
STS-GCN	Space-time-separable Graph Convolutional Network
SVM	Support Vector Machine
SWS	Slow-wave Sleep
TCN	Temporal Convolutional Network
TN	True Negative
TP	True Positive
TPR	True Positive Rate

Acknowledgment

It takes a village to raise a child is a well-known proverb, and I strongly believe that completing a Ph.D. also requires the support of a village. Therefore, I would like to express my sincere gratitude to the people who played significant roles in my Ph.D. journey.

First and foremost, I would like to express my deepest gratitude to my advisor, Professor ZHAO Qiangfu, for his unwavering support, invaluable guidance, and continuous encouragement during the entirety of my doctoral journey. His expertise and insightful feedback played an integral role in guiding the direction of my research, completing the complexity of doctoral research, and writing this thesis. During the third year of my doctoral research, Professor ZHAO Qiangfu generously offered me professional support, which was made possible through our biweekly meetings and numerous email exchanges. I greatly appreciate our bi-weekly meetings, which not only serve as crucial checkpoints to keep me on track academically, but also offer me with plenty of encouragement. Whenever I complete a research project and finish a journal manuscript, Professor ZHAO Qiangfu exhibits exceptional patience and consistently devotes his time to correcting and guiding me. His dedication ensures that my manuscript is seamlessly aligned with the standard requirements. Furthermore, in instances where my submitted manuscript is rejected, Professor ZHAO Qiangfu not only offers encouragement that boosts my confidence, but also provides me with valuable revision suggestions. These suggestions serve to improve my professional experience and growth. To be quite frank, Professor ZHAO Qiangfu has bestowed upon me the invaluable lessons of fostering consistency and concentration in my research pursuits. His guidance has enabled me to grasp the essence of being a researcher. Words cannot adequately convey the depth of my gratitude.

Second, I would like to extend my heartfelt appreciation to Professor CHENG Zixue from the University of Aizu for his unwavering assistance and support throughout my master's program and the initial two years of my doctoral journey at the University of Aizu. His guidance has illuminated my path, and their insights have been instrumental in helping me navigate the intricate landscape of sleep research, experimental setup, and methodologies. In each discussion or weekly meeting, Professor CHENG Zixue consistently presents captivating topics and offers perceptive viewpoints, which can continually motivate me to strive for continuous improvement.

Third, I am also deeply appreciative of the contributions made by the members of my thesis committee: Professor LIU Yong, Professor ZHU Xin, and Professor Michael Cohen from the University of Aizu, for their constructive critiques and valuable suggestions that have significantly enriched the quality of the final output of my research. In particular, Professor LIU Yong took a lot of time and energy to revise my second Journal thesis. Thank you for your time and effort in reading my second thesis thesis for providing valuable feedback and thought-provoking questions. Your insights have

greatly enriched the quality of my work.

In addition to the above professors, I am indebted to my exceptional lab mates and every friend and professor from the University of Aizu, whose support has been a constant source of motivation. Our informal chats, whether conducted through screens during lockdowns or in person whenever circumstances allow, serve as a continuous source of motivation and inspiration, propelling me forward even during the most challenging times. I am proud to say that we became more than just lab partners, but good friends. LU Chenghong, thanks for always providing a steady supply of snacks that brought much-needed sweetness to our intense work sessions. WANG Zhishang, your late-night conversations help me to keep my self-doubt in check, and your suggestion and discussion in the process of my research have enormously enriched me personally. CHEN Hongbo, your exceptional academic expertise and positive feedback have played a pivotal role in refining my academic writing and sharpening my arguments. My roommate LIANG Yuxiao, your enthusiasm and upbeat character made spending long hours in the housing an enjoyable experience. Your affection and concern have served as powerful motivators, propelling me to push myself further and evolve into a better version of myself. I would also like to extend my gratitude to the University of Aizu; it has been a privilege to be a part of this institution, and the memories forged here will forever hold a special place in my heart.

Lastly, I want to express my deepest gratitude to my family who believe in my abilities and support. Thank you for always being my rock, accompanying me through the highs and lows of this academic journey. Your encouragement played an integral role in my accomplishments. To my mom and dad: Thank you for everything. I dedicate this Ph.D. thesis to you. As I bring my three-year-long Ph.D. journey to a close, I can genuinely say that I am immensely proud of my accomplishments. This fulfilling journey has not only shaped me as a researcher but also as an individual, imparting the invaluable lesson that perseverance yields fruitful rewards in the long run.

LI Menglei,
April 2024,
Aizuwakamatsu, Japan

Abstract

Sleep is crucial for both physical and psychological well-being. However, an increasing number of modern individuals are affected by sleep disorders, which have become a widespread societal issue. Insomnia, sleep apnea, and restless leg syndrome are common sleep disorders that can lead to difficulty falling asleep, maintaining sleep, or achieving restorative deep sleep. sleep stage classification is used as an aid in the diagnosis and treatment of sleep disorders, while polysomnography is considered the gold standard for sleep stage classification that assesses sleep by simultaneously monitoring multiple physiological signals, like electrooculogram (EOG) and electroencephalogram (EEG). Traditionally, sleep stage classification has relied on labor-intensive manual scoring or limited-channel polysomnography. Namely, sleep stage classification historically relies on the subjective judgment of sleep experts. Therefore, different sleep experts may have slight variations in their interpretation of the same data, leading to some degree of subjectivity in sleep stage classification.

To mitigate subjectivity and improve consistency in sleep stage classification, automatic sleep stage classification algorithms have been developed that objectively analyze bioelectric signals and classify sleep stages, thereby reducing subjectivity in manual scoring by sleep experts. However, these bioelectric signals are non-Euclidean graph-structured data. Due to their exceptional processing of graph-structured data, graph neural networks (GNNs) are widely used for automatic sleep stage classification, yielding significant results.

However, there are two major deficiencies of existing GNN-based methods for sleep stage classification using single-channel EEG data. First, although GNNs are powerful tools for analyzing graph-structured data, they typically rely on a static adjacency matrix that may not fully capture the spatial information and relationships between each EEG channel (electrode). Second, the importance of spatiotemporal relationships in classifying sleep stages based on EEG data is overlooked. EEG signals are not only spatially distributed across electrodes, but also vary over time as individuals transition through different sleep stages. In our first work, we propose a combination of a dynamic and static spatiotemporal graph convolutional network (ST-GCN) with inter-temporal attention blocks based on EEG to overcome two shortcomings. Specifically, we leverage spatial graph convolutions and temporal convolutions to effectively model EEG data. To capture the enriched global context and topology, we use a combination of dynamic and static ST-GCN. We also use temporal convolutions with dilation to expand the temporal receptive field and effectively capture long-range temporal dependencies in EEG signals, which is critical for accurate sleep stage classification. Notably, we introduce attention blocks for the first time in the field of sleep stage classification. The intertemporal attention blocks allow us to model the relationships between different EEG channels, thereby capturing long-range dependencies that help the model understand how EEG signals at different time points influence each other, which is essential

for accurate classification. Our proposed model for sleep stage classification based on EEG data demonstrates better performance compared to some other state-of-the-art models.

Despite the fact that our proposed single-channel EEG-based model has provided better classification accuracy, the complementary nature of multimodal electrophysiological signal characteristics is overlooked. The existing multi-stream sleep staging network relies predominantly on EOG and EEG signals as its primary inputs, adeptly amalgamating the extracted multimodal features cleverly merged to improve performance. Moreover, according to our observation, few researchers have focused on the motor information of electrophysiological signals in the context of sleep stage classification. This motor information can provide valuable information about sleep stages and improve the accuracy of sleep stage classification. In addition, the problems of overparameterization and suboptimal classification accuracy are common challenges in classification tasks based on Deep Learning, especially when applied to complex tasks such as sleep stage classification. To address the above challenges, in our second proposed work, we develop an efficient graph-based multi-stream model called 4s-SleepGCN that merges EEG, EOG, and the corresponding motion information into a unified multi-stream network framework for sleep stage classification. In our proposal, the EEG signal, EOG signal, and corresponding motion information are each fed separately into the single-stream model. In each single-stream model, the positional relationship of the modal sequences within a recording is first considered by position embedding. Position embedding can help our proposed models better capture the sequential dependencies and temporal context in different signals, thereby improving the feature representation for sleep stage classification. Building upon this foundation we use graph convolution to capture spatial features and employ temporal convolution at multiple scales to capture temporal dynamics and extract more discriminative contextual temporal features. Finally, the prediction of sleep stage classification is calculated by the weighted summation method of the four softmax scores. Our proposed 4s-SleepGCN demonstrates exceptional performance in sleep stage classification when compared to existing state-of-the-art methods. Moreover, our single-stream model is notably lightweight and demands fewer parameters. It can be proved that our proposed single-stream baseline can be introduced as a strong and powerful baseline for sleep stage classification. Therefore, the two models we have presented stand as effective tools that can assist sleep experts in assessing sleep quality and diagnosing sleep-related disorders.

Chapter 1

Introduction

Sleep is a fascinating field of research with implications for a wide range of disciplines, from neuroscience to psychology and beyond. It is well known that good sleep is essential for cognitive function, memory consolidation, emotional regulation, and overall health. To understand and analyze sleep patterns, researchers and clinicians employ a specialized field known as sleep medicine, which relies on various techniques to monitor and classify the different sleep stages that individuals undergo during the night. One of the key challenges in sleep medicine is accurately classifying these sleep stages, commonly referred to as the sleep stage classification problem. The sleep stage classification problem involves the complex task of categorizing different sleep stages based on recorded data from a variety of sources, including electroencephalography (EEG) to measure brain wave activity, electrooculography (EOG) to monitor eye movements, electromyography (EMG) to assess muscle tone, and additional physiological signals such as heart rate and respiratory rate. These sleep stages typically include wakefulness, rapid eye movement sleep (REM), and several non-rapid eye movement (NREM) stages, each characterized by unique patterns of physiological activity. To truly comprehend the profound impact of sleep on our lives and address sleep-related issues effectively, it is imperative that these sleep stages need to be accurately classified and analyzed. This necessity underpins the significance of the sleep stage classification problem.

1. **Understanding Sleep Patterns**: Sleep is not a uniform state; rather, it comprises a continuum of stages that transition throughout the night. Accurate sleep stage classification provides essential insights into the temporal organization of sleep, including the duration and sequencing of each stage. This understanding is foundational for assessing sleep quality and identifying abnormalities.

2. **Diagnosing Sleep Disorders**: Sleep disorders affect millions of individuals worldwide, ranging from common conditions like insomnia and sleep apnea to more rare disorders such as narcolepsy. Accurate classification of sleep stages is crucial for diagnosing these disorders, as each often exhibits distinct deviations from normal sleep patterns. For example, obstructive sleep apnea is characterized by recurrent interruptions in breathing during sleep, predominantly occurring during specific stages, notably REM sleep.

3. **Personalized Treatment Plans**: The diverse nature of sleep disorders and the variations in individual sleep architecture demand personalized treatment approaches. By precisely classifying sleep stages, clinicians can tailor interventions to address specific issues that arise during particular stages of the sleep cycle. This personalized approach enhances treatment efficacy and patient outcomes.

4. **Health and Well-Being**: Quality sleep is vital for physical and mental health. Accurate sleep stage classification helps researchers and healthcare professionals explore the intricate relationships between sleep patterns and various health outcomes. This knowledge can inform strategies for improving overall well-being and reducing the risk of chronic health conditions associated with sleep disturbances.

In view of these critical reasons, the need for precise sleep stage classification cannot be overstated. It forms the cornerstone of advancements in sleep medicine and allows for better diagnosis, treatment, and research in the quest for healthier, more restorative sleep for individuals worldwide. The realm of sleep stage classification has witnessed significant advancements in recent years, driven by the convergence of technology, data

analytics, and a growing awareness of the importance of sleep in our daily lives. Historically, sleep stage classification relies heavily on traditional polysomnography (PSG), which includes a combination of EEG, EOG, EMG, and other physiological signals. This approach lays the foundation for sleep research and diagnosis. It provides valuable insight into the different characteristics of each sleep stage and serves as a benchmark for subsequent methods. Manual scoring of sleep stages by trained experts has been the gold standard for many years. However, this approach is labor-intensive, time-consuming, and subject to inter-scorer variability. Studies have explored the limitations of manual scoring, highlighting the need for more automated and consistent methods. In recent years, machine learning and deep learning techniques have been increasingly used to classify sleep stages. These approaches leverage large datasets to train algorithms that can automatically classify sleep stages with a high degree of accuracy. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid architectures have shown promise in improving classification performance. In addition, graph neural networks (GNNs) have been successfully applied to the processing of non-Euclidean data such as EEG and EOG. Despite the progress made, challenges persist in sleep stage classification. Variability in sleep patterns across individuals, the need for powerful models, and the development of reliable monitoring methods are among the ongoing research areas.

1.1 Sleep Stage Classification Utilizing Bioelectrical Signals: Motivation

Sleep stage classification utilizing bioelectrical signals is a significant area of research within the fields of sleep medicine, neuroscience, and biomedical engineering. This problem involves the analysis and interpretation of various physiological signals, primarily EEG and EOG, to determine the different stages of sleep a person is in. There are three main motivations that inspired us and are reflected in this dissertation as follows:

- **Motivations #1: The physiological signals are non-Euclidean data, which means that the traditional methods for analyzing Euclidean data, like images and structured tabular data, are not directly applicable to physiological signals.**

  EEG or EOG data are characterized by their temporal and spatial complexity. These signals are recorded over time from multiple electrodes placed at various locations on the scalp (EEG) or around the eyes (EOG), resulting in multivariate time-series data with inherent spatial relationships. The interconnections among electrodes are irregular, forming a graph structure rather than a regular grid. The non-Euclidean nature of EEG and EOG data presents challenges when applying traditional machine learning methods that assume grid-like structures, such as regular images. Graph-based approaches, such as graph convolutional networks (GCNs) [1–3], are particularly well-suited for dealing with non-Euclidean data. Thus, our goal is to use a GCN framework that can achieve state-of-the-art performance in sleep stage classification utilizing bioelectrical signals such as EEG and EOG data.

- **Motivations #2: Enhancing sleep staging efficiency to improve the diagnosis of sleep disorder.**

  As mentioned earlier, sleep disorders, which affect a significant portion of the global population, have far-reaching implications for health, productivity, and overall well-being. Achieving an accurate and efficient sleep stage classification model is essential for diagnosing these disorders and providing appropriate treatment. While traditional manual methods of sleep stage classification have been considered the gold standard, advancements in technology and computational methods offer the potential for more objective, reliable, and scalable approaches. Our objective is to explore and refine automated techniques of sleep stage classification using deep learning and signal processing algorithms. By enhancing the accuracy of sleep stage classification and mitigating the inherent subjectivity in manual scoring, this endeavor can contribute to the optimization of

diagnosis and treatment of sleep disorders, ultimately improving the sleep quality of individuals and long-term health outcomes.

- **Motivations #3: Different modalities of physiological signals have different contributions to sleep stage classification with distinct impacts on different sleep stages.**

  To achieve a more accurate classification of sleep stages, analysis of multiple bioelectrical signals is usually required rather than relying solely on a single signal. Sleep is a complex physiological phenomenon with various sleep stages that involve distinct changes in brain activity, eye movements, muscle tone, and more. These changes are best captured and understood by considering a combination of different signals. Physiological signals such as EEG and EOG offer insights into distinct aspects of sleep-related phenomena, each of which has a unique influence on different sleep stages. By comprehensively exploring the individual and collective contributions of these signal modalities to sleep stages classification, we can shed light on their specific impacts on the delineation of wakefulness, REM sleep, and NREM stages. In order to enhance the accuracy and depth of sleep stage classification methods, we delve into the nuanced interactions between these signals and sleep stages. Consequently, this approach to combining multi-stream biological signal features can advance our understanding of sleep architecture and its implications for health and well-being.

1.2 Research Goals

The foremost goal of this study is to advance the state-of-the-art in sleep stage classification. We aim to develop and evaluate novel deep learning-based models that can provide a more accurate and reliable classification of sleep stages. This includes exploring innovative feature extraction techniques and model architectures. In addition to traditional polysomnography, our study explores the potential of multimodals, including motion information from EEG and EOG. We aim to assess the feasibility and

accuracy of utilizing multimodal for sleep stage classification, with the goal of enabling more accessible and cost-effective monitoring solutions.

1.3 Dissertation Outline

The dissertation is primarily divided into six chapters: Introduction, Related works, A signal-channel EEG-based approach: **An attention-guided spatiotemporal graph convolutional network for sleep stage classification**, A multimodal physiological signals-based approach: **4s-SleepGCN**, and conclusion and future work. The dissertation outline is depicted in Figure 1.1.

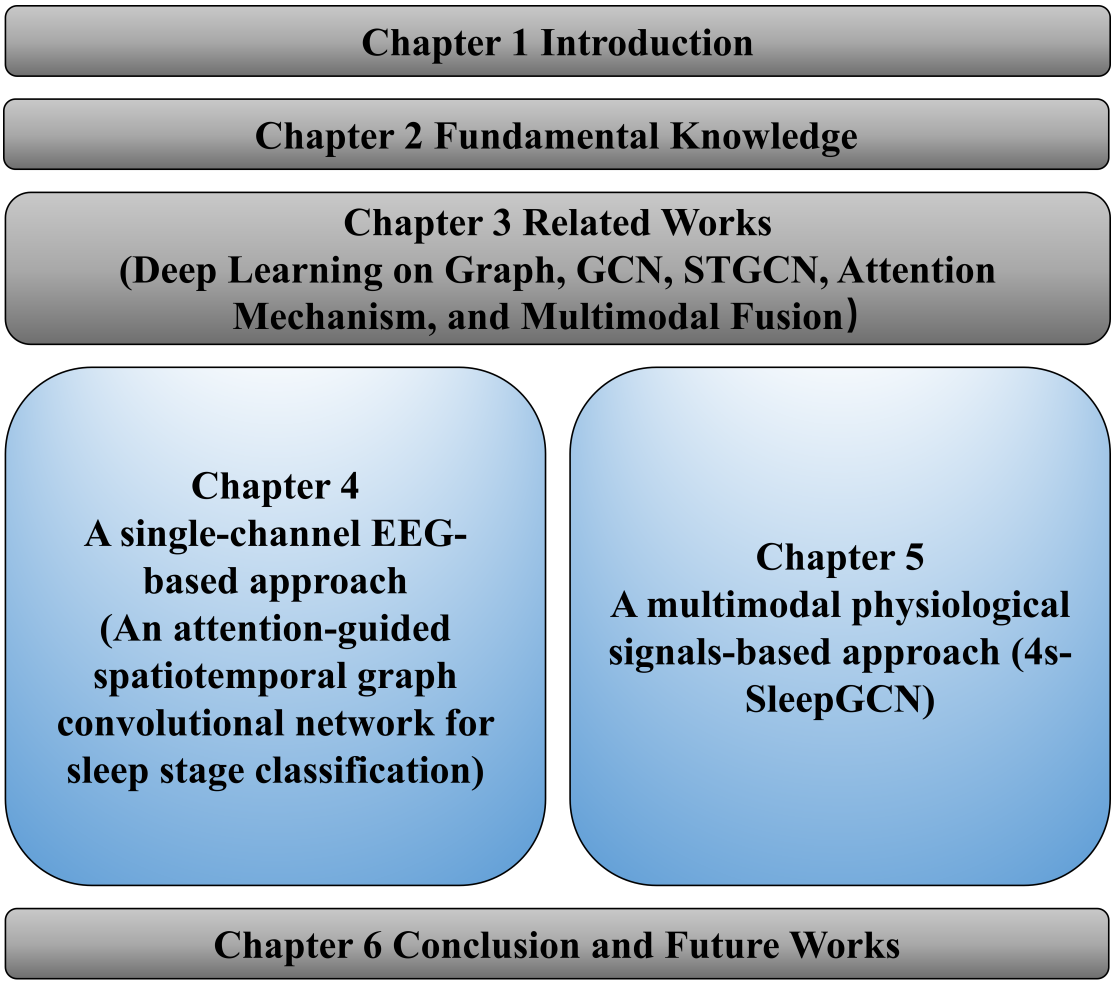


Figure 1.1: **The outline of the dissertation.**

Chapter 1 serves as the introductory section of this dissertation, with the primary goal of establishing the research background and elucidating its significance. In addi-

tion, we briefly introduce the main results obtained so far in the literature. Finally, the main motivations and contributions are presented.

In Chapter 2 of Fundamental Knowledge, we introduce the concept of sleep and provide a brief overview of its effects on health and well-being. Moving forward, this chapter proceeds to outline prevalent sleep disorders, offering a comprehensive understanding of their characteristics and manifestations. Additionally, we expound upon the repercussions of these sleep disorders on individuals' lives. Furthermore, we provide an introduction to PSG as a comprehensive technique for monitoring sleep and its various physiological parameters. We then provide an overview of sleep stage scoring manuals and standardized criteria based on PSG data. In addition, we outline the unique characteristics associated with each sleep stage according to the latest standardized criteria.

Chapter 3 provides an overview of pertinent studies and techniques employed in this work. In the first two sections of this chapter, we delve into the historical evolution of GCNs and Spatial-Temporal Graph Convolutional Networks (ST-GCN) and describe the basics of the two models, shedding light on their respective advancements and milestones. Also, we highlight important work or studies that led to and contributed to the development of GCNs and ST-GCN and explore a wide range of applications where the two frameworks have been successfully applied, with specific examples (e.g., sleep stage classification, traffic prediction, climate analysis). Third, we provide an overview of attention mechanisms and their significance in enhancing the capabilities of machine learning models. Recent advances and applications in the area of attention mechanisms are presented in this section of attention mechanisms. Finally, we not only introduce the concept of multimodal fusion and its importance in combining information from different modalities to improve understanding and performance but also show the role of multimodal fusion in various applications.

In Chapter 4, a novel model for sleep stage classification using single-channel EEG data is proposed. The proposed model is based on the architecture of the ST-GCN, which can effectively capture the global context-enriched topology and employs temporal convolution with dilation to enlarge the temporal receptive field. Moreover, to

the best of our knowledge, this is the first time that an intertemporal attentional block has been integrated into a sleep stage classification network. The introduction of the attention blocks models the relationship between different EEG channels to capture long-range dependencies for sleep stage classification, thereby improving the ability of our proposed sleep stage classification model.

Chapter 5 underscores the significance of employing multimodal physiological signals in sleep stage classification. In Chapter 4, a novel graph-based multi-stream fusion model called 4s-SleepGCN for automatic sleep stage classification is presented. This proposal can better utilize the classification performance by embedding the positional relationship of the modal sequence. In addition, the EEG data, EOG data, and corresponding motion information are fused into a unified multi-stream network framework to improve the proposed model's ability to differentiate between different sleep stages. Our newly proposed 4s-SleepGCN model introduces the incorporation of EEG and EOG motion information, marking a pioneering step in sleep stage classification. This innovation represents a significant advance in the field of sleep stage classification research.

In Chapter 6 of our dissertation, we highlight primarily the contributions made throughout the research journey. In particular, we summarize the main findings, innovations, and insights that emerged from this work. In addition, this chapter provides a critical reflection on the significance and implications of the study's findings, methodology, and overall approach. Moreover, the chapter focuses on a thorough discussion of our proposed sleep stage classification model, including its strengths and limitations, while also outlining a roadmap for future research plans to improve and refine the performance of the sleep stage classification model. This chapter is pivotal in terms of demonstrating the long-term impact and potential continuation of the research beyond the current study.

1.4 Thesis Objectives and Contributions

The dissertation is a report on my three years of research on GCN architecture and framework design for the sleep stage classification task. To efficiently categorize distinct sleep stages, we present two GCN-based approaches within this study: (1) a single-channel EEG-based approach and (2) a multimodal physiological signals-based approach. Chapters 3 and 4 present our two proposed novel approaches and experimental results. First, as reported in Chapter 3, we propose a combination of dynamic and static ST-GCN, augmented by inter-temporal attention blocks. This proposal is formulated with the aim of automating sleep stage classification using single-channel EEG data. Second, in Chapter 4, considering the complementary potential of PSG signals, we introduce a highly efficient graph-based multi-stream model termed 4s-SleepGCN. This innovative model fuses features extracted from EEG, EOG, and the corresponding motion information to improve the precision of sleep stage classification. The dissertation is approached with the goal of leveraging the GCN framework to process non-Euclidean data, such as physiological signals. The overall goal is to achieve a level of performance in sleep stage classification that surpasses that of its state-of-the-art counterparts. The main contributions to our proposed approaches can be summarized as follows:

1. Our proposed single-channel EEG-based approach (An Attention-guided Spatiotemporal Graph Convolutional Network for Sleep Stage Classification).

    - In previous work, sleep stage classification is achieved by complex modeling. In contrast, our proposed method is to leverage spatial graph convolutions along with interleaving temporal convolutions to achieve spatiotemporal modeling, which can be simpler yet more efficient.
    - The inter-temporal attention blocks are introduced to achieve an automatic sleep stage classification, which can withdraw the most informative information across space and time, further proving that capturing spatiotemporal relation plays an important role in sleep stage classification.
    - The proposed model significantly outperforms state-of-the-art methods on

the sleep-EDF and the subgroup III of the ISRUC-SLEEP dataset. Our proposed method achieves better performance with 91.0% and 87.4% accuracy, both outperforming the state-of-the-art methods (86.4% and 82.1%).

2. Our proposed multi-modal physiological signals-based approach (4s-SleepGCN: Four-Stream Graph Convolutional Networks for Sleep Stage Classification).

    - To the best of our knowledge, we are the first to utilize a multi-stream fusion strategy to facilitate the fusion of EEG signals, EOG signals, and the corresponding motion stream, which significantly outperforms the state-of-the-art methods on two benchmark datasets for sleep stage classification. Furthermore, the motion modality is shown to be a beneficial addition to sleep staging.
    - In each single-stream model, we utilize the position embedding method along with spatial-temporal convolutions to model spatial-temporal relationships effectively and classify sleep stages.
    - We propose a lightweight, single-stream solid baseline that is more potent than most previous methods. We hope that the solid baseline will be helpful for the study of automatic sleep stage classification.
    - On the Sleep-EDF-39 and Sleep-EDF-153 datasets, our proposed model named 4s-SleepGCN outperforms both single-stream and two-stream models. The experimental results underscore the importance of multiple information. Our proposed model addresses the current deficiencies of multi-modal learning in sleep staging, paving the way for multi-modal learning in sleep stage classification.

In this work, we develop and rigorously evaluate a set of novel GCN-based models tailored for sleep stage classification. These models exhibit state-of-the-art performance, surpassing existing methodologies in accuracy and robustness. Our comprehensive analysis demonstrates the potential for a more reliable and precise classification of sleep stages. To assess the generalizability of our models, we conducted extensive

validation across diverse datasets. Our findings showcase the adaptability of our approaches to different data sources, further underscoring their practicality. The results collectively underscore the contributions of this study to the ever-evolving landscape of sleep science and technology. By advancing the accuracy and practicality of sleep stage classification, we aim to catalyze further progress in understanding the critical role of sleep in human health and to provide valuable tools for clinicians and researchers alike.

Chapter 2

Fundamental Knowledge

2.1 What is Sleep?

Sleep is a naturally recurring state of rest that is essential for the proper functioning and well-being of living organisms, including humans [4,5]. It is a fundamental physiological process that allows the body and mind to rest, recharge, and rejuvenate [6]. The one activity we spend most of our life doing is sleep. The average person spends about 26 years sleeping in their life which equates to 9,490 days or 227,760 hours. Thereby underscoring the pivotal importance of achieving restful and sufficient slumber as an integral constituent of a holistic and healthful way of life. During sleep, the body undergoes a variety of regenerative processes, including tissue repair, muscle growth, and memory consolidation [7]. These processes are essential for physical and cognitive health. For example, good-quality sleep not only regulates metabolism and contributes to cardiovascular health, but also promotes mental clarity, concentration, and the consolidation of memories, as shown in Figure 2.1.

In addition, sleep plays a central role in emotional well-being and stress reduction. It provides the brain with the opportunity to process information, improve cognitive function, and regulate emotions. Studies have consistently shown that people who consistently get enough high-quality sleep tend to experience better mental health and resilience [8,9]. In addition to its immediate benefits, sleep also has a profound impact on

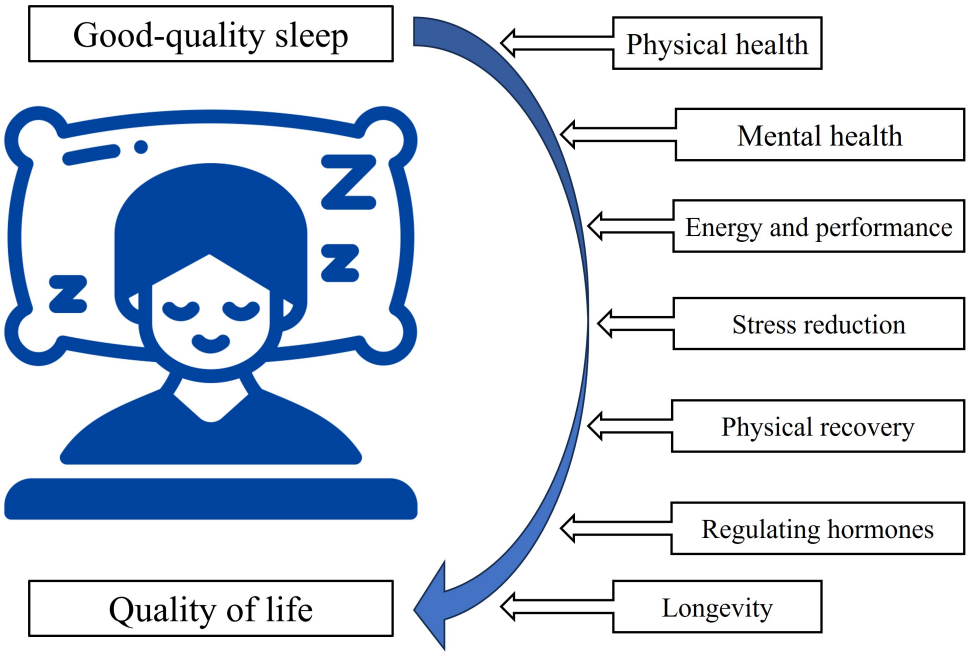


Figure 2.1: **The relationship between good-quality sleep and the quality of life.**

long-term health and longevity. Numerous studies have established a strong association between sufficient sleep and longer life expectancy [10–12].

In summary, sleep is not just a passive state of rest, but a dynamic and essential process that contributes to a robust and satisfying life. Healthy sleep habits can effectively enhance general well-being, cognitive abilities, and emotional stability, equipping individuals with the capacity to lead more dynamic and fruitful lifestyles. However, it is important to acknowledge that sleep disorders can significantly undermine these advantages. Therefore, prioritizing and nurturing healthy sleep habits is crucial for overall health and longevity.

2.2 Sleep Disorders

In today's rapidly evolving global environment, characterized by intense social competition, escalating work-related pressures, and an increasingly aging demographic, the prevalence and impact of sleep disorders have emerged as significant public health concerns [13–15]. Sleep disorders encompass a spectrum of conditions that impede an individual's ability to achieve adequate and restorative sleep. Primarily, these disor-

ders manifest in forms such as insomnia, circadian rhythm disturbances, and obstructive sleep apnea syndrome [16, 17]. Simultaneously, sleep disorders have escalated in prevalence among adults. Numerous unfavorable factors may exert a dual influence, contributing to the emergence of sleep disorders while also being intensified by them, such as the consumption of substances like caffeine, nicotine, and alcohol [18], sleep habits [19], and comorbid diseases [20]. According to the American Sleep Association (ASA), roughly 50 to 70 million adults in the United States of America experience sleep disorders [21]. In addition, sleep apnea is estimated to affect a significant portion of the population, with prevalence rates ranging from 2% to 4% among adults and 1% to 3% among children [22–24]. As outlined in the newly published third edition of the International Classification of Sleep Disorders (ICSD) [25], sleep disorders can be classified into seven major categories. These are respectively insomnia disorders, sleep-related breathing disorders, central disorders of hypersomnolence, circadian rhythm sleep-wake disorders, sleep-related movement disorders, parasomnias, and other sleep disorders, as shown in Figure 2.2. Here's a concise overview of these sleep disorders as follow:

- **Insomnia** stands as a common sleep disorder, distinguished by challenges in falling asleep, staying asleep, or experiencing non-restorative sleep, even in the presence of suitable conditions and surroundings for slumber. As a result, people experience daytime impairment due to insomnia, as evidenced by fatigue, heightened irritability, diminished focus, and a general decrease in overall well-being.

- **Sleep-related breathing disorders** are a group of conditions that involve disruptions in a person's breathing patterns during sleep, which can lead to various health problems, such as reduced oxygen intake, fragmented sleep, and daytime sleepiness. In particular, Obstructive Sleep Apnea (OSA) and Central Sleep Apnea (CSA) are two prominent and distinct examples of sleep-related breathing disorders.

- **Central Disorders of Hypersomnolence (CDH)** represents a grouping of sleep

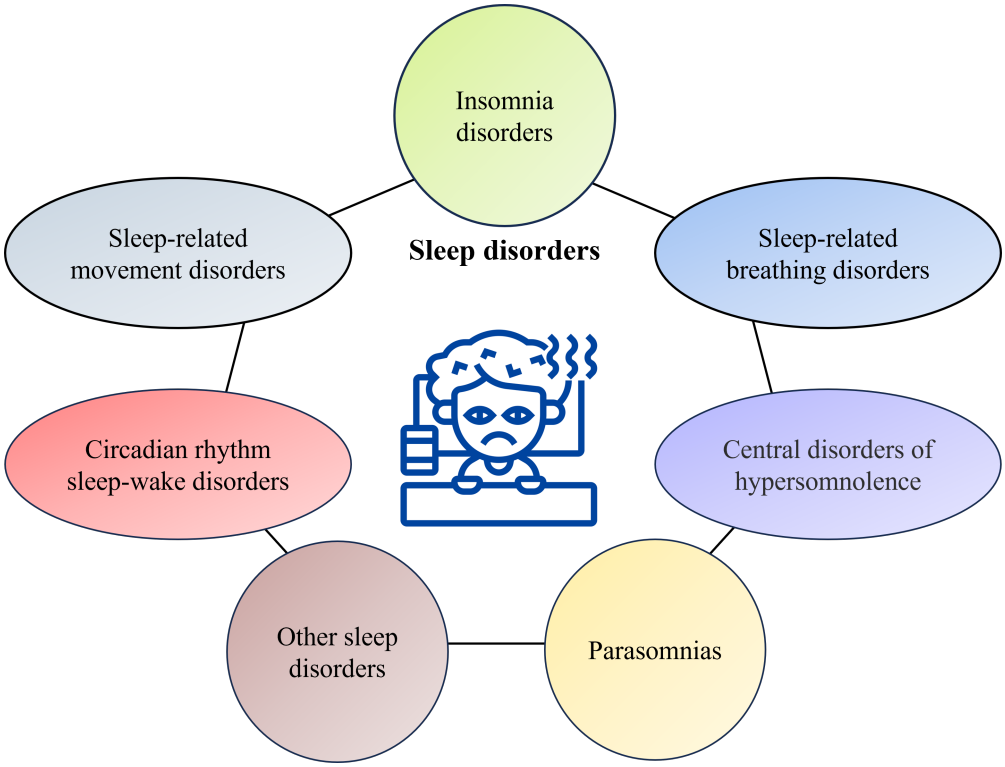


Figure 2.2: **The classification of sleep disorders based on the third edition of ICSD.**

disorders marked by excessive daytime sleepiness, even when sufficient sleep is obtained during the night. These disorders are generally caused by dysfunction in the central nervous system's regulation of wakefulness and sleep. Therein, Narcolepsy is one of the most well-known CDHs.

- **Circadian rhythm sleep-wake disorders** also known as circadian rhythm sleep disorders, are a group of sleep disorders characterized by disruptions in the natural timing of sleep and wakefulness that result in difficulty falling asleep, staying awake, or maintaining a regular sleep schedule that affects the timing of sleep.

- **Sleep-related Movement Disorders (SRMDs)** are conditions in which a person experiences involuntary movements during sleep, disrupting the sleep cycle and leading to sleep disturbances. The movements that occur can vary in intensity and type and often lead to daytime fatigue and impaired overall functioning. Moreover, the most common SRMDs are Restless Legs Syndrome (RLS) and Periodic Limb Movement Disorder (PLMD).

- **Parasomnia** serves as a category of sleep disorders in which people experience abnormal movements, behaviors, emotions, perceptions, and dreams during different stages of sleep. Parasomnias are classified primarily by the sleep stages in which they occur into non-rapid eye movement-related parasomnias, rapid eye movement-related parasomnias, and others.

As we know, sleep disorders profoundly impact an individual's well-being, affecting not only their physical health but also their mental and emotional state. Proper diagnosis, comprehensive comprehension, and effective treatment are crucial to addressing these sleep disorders and improving the quality of life for individuals grappling with their effects. Hence, sleep staging plays a pivotal role in the realm of sleep disorder diagnosis and management. It can provide essential information that assists healthcare professionals in formulating accurate diagnoses and personalized treatment strategies. This is particularly important given the varied incidences and characteristics of sleep disorders across different sleep stages. To visually illustrate the relationship between sleep disorders and different sleep stages, we present a selection of common sleep disorders categorized by their incidence during different sleep stages in Table 2.1. Therein, these common sleep disorders include insomnia [26], sleep apnea [27], narcolepsy [28], sleepwalking [29], night terrors [30], Rapid eye movement sleep behavior disorder (RBD) [31]. Sleep disorders exhibit diverse manifestations across the different stages of sleep, each of which is characterized by unique physiological and neurological attributes. Consequently, the nature and characteristics of sleep disorders can significantly vary, contingent upon the specific sleep stage impacted.

In summary, sleep stage classification offers a detailed and objective assessment of a person's sleep architecture. By examining the distribution of sleep stages throughout the night, sleep specialists can make well-informed judgments regarding diagnosis, treatment, and overall sleep management. This approach significantly enhances our ability to address sleep-related disorders and foster healthier sleep patterns, thereby improving individual well-being. Consequently, the practice of monitoring and analyzing sleep stages in accordance with the various sleep stages observed throughout the night

Table 2.1: The relationship between sleep disorders and different sleep stages.

Disorder	**State of occurrence**	**Essential Features**
Insomnia	Insomnia can impact all sleep stages and often become more pronounced during the light sleep stages (N_1 and N_2 stages).	Difficulty falling asleep; Staying asleep; Waking up too early.
Sleep Apnea	Sleep apnea is often more pronounced during the deeper sleep stages (N_3 and REM stages).	Interruptions in breathing; Frequent awakenings.
Narcolepsy	Narcolepsy involves rapid transitions between wakefulness and REM stage.	Daytime sleepiness; Sleep attacks; Disruptions in REM stage.
Sleepwalking	Sleepwalking typically occurs during the N_3 stage.	Walking or complex behaviors during sleep.
Night Terrors	Night Terrors occur the NREM stage (the transition from the N_3 stage to the N_2 stage).	Intense fear and panic.
RBD	RBD occurs during REM stage.	Loss of muscle atonia.

[1] RBD is an abbreviation for REM sleep behavior disorder.
[2] NREM and REM are the abbreviations for Non-rapid Eye Movement and Rapid Eye Movement, respectively.

is of great significance.

2.3 Sleep Staging

Sleep staging, also known as sleep classification or sleep scoring, is a pivotal component of sleep science that provides a profound insight into the intricate phases that our minds and bodies undergo during slumber. Within the realm of sleep staging, sleep is classified into distinct stages according to characteristic patterns of brain activity, eye movement, muscle tone, and other physiological parameters. Numerous studies [32–35] have consistently established that sleep stage scoring stands as the gold standard for the comprehensive analysis of human sleep. In order to obtain an accurate sleep staging score, the recognition of different sleep stages relies on various physiological measurements obtained from a polysomnography (PSG) sleep study. As a result, the analysis of PSG data is considered a representative criterion for sleep stage scoring [36].

2.3.1 Polysomnography (PSG)

Polysomnography (PSG) [37], a cornerstone diagnostic modality in sleep medicine, is employed for the in-depth analysis of physiological parameters during sleep. PSG functions by recording a spectrum of bodily functions through sensors attached to various body parts as an individual sleeps or attempts to sleep. It encompasses the measurement of brain activity (via electroencephalography, EEG), eye movements (electrooculography, EOG), muscle activity (electromyography, EMG), cardiac rhythm (electrocardiography, ECG), as well as respiratory effort, airflow, and blood oxygen saturation. This comprehensive approach allows for a detailed understanding of sleep patterns and anomalies, thereby facilitating accurate diagnoses and effective treatment plans for sleep-related disorders. The origins of the contemporary PSG can be attributed to the pioneering work of Caton on 4^{th} August 1875 [38]. Caton's research unveiled brain wave activity in the mammalians, marking a crucial advancement in the field. As time continued to pass, the concept of recording the electrical activity of human beings grad-

ually come to light and became recognizable to the general public, with its origins tracing back to the late 19^{th} and early 20^{th} centuries. The German psychiatrist Hans Berger [39] discovered the EEG in the 1920s, marking a significant milestone in the understanding of brain activity. Berger's contributions have had a substantial impact on the application of EEG in the fields of medicine and science. The development of PSG into a comprehensive tool for monitoring sleep-related physiological activities is a process that requires time and progress across various domains. In the mid-20^{th} century, technological advancements, with the advent of innovations such as amplifiers, recording equipment, and signal processing techniques, play a pivotal role in the refinement of PSG techniques. With the continuous improvement of these technologies, the capability to simultaneously record multiple physiological signals (i.e., EOG, EMG, ECG) becomes feasible. Accordingly, PSG has evolved from a research tool to a clinical diagnostic tool. Namely, sleep specialists can glean valuable insights into an array of distinct sleep disorders by analyzing the intricate interactions among various physiological parameters during sleep [37, 40]. Especially in the diagnosis of sleep disorders like sleep apnoea, due to the fact that PSG can accurately monitor and analyze irregularities in breathing and oxygen levels. In summary, with the advancement of digital technology and data analysis techniques, PSG has become more accessible and standardized. This allowed the general public to become more aware of the importance of sleep monitoring and sleep disorders.

As illustrated in Figure 2.3, PSG involves the use of attached electrodes and various sensors placed on different parts of the body to record and monitor a wide range of physiological parameters during sleep [41]. Therein, a total of six EEG electrodes are employed for the collection of EEG recordings (C_3, C_4, M_1, M_2, REF, and GND, respectively). Furthermore, the EOG is utilized to capture eye movements, while the EMG of the chin is employed to monitor facial muscle tone. Additionally, the PSG recordings contain a one-channel ECG and an EMG recording obtained from the tibialis anterior muscle. The effort sensors are used to monitor the thorax and abdominal respiratory expansions. Moreover, respiratory events could be assessed using a nasal can-

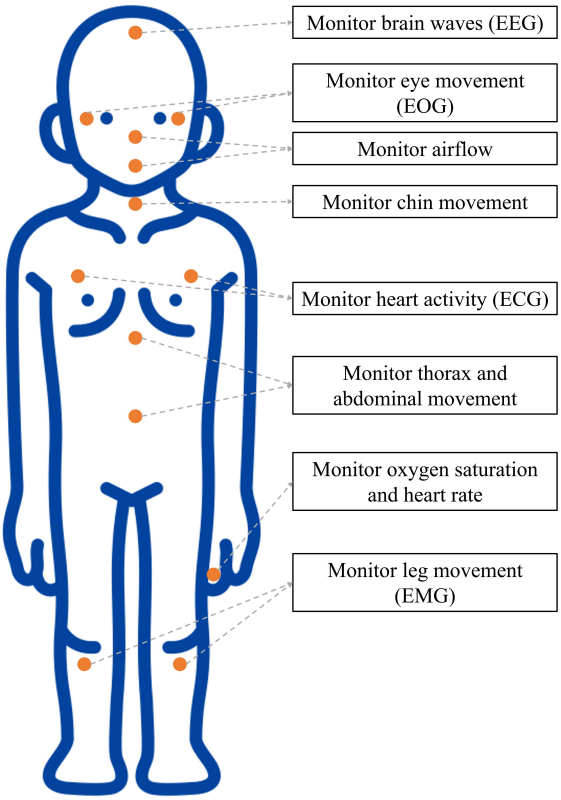


Figure 2.3: **PSG recordings consist of electrodes placed on the head, face, chest, hand, and legs, including the EEG, EOG, chin EMG, airflow, ECG, pulse oximetry, respiratory effort (thoracic/abdominal), snore microphone, and body position sensor.**

nula functioning as a pressure sensor, while occurrences of snoring could be captured through a microphone positioned adjacent to the larynx. Capillary oxygen saturation is monitored using light-sensitive finger-pulse oximetry. These sensors work together to provide a comprehensive picture of an individual's sleep patterns and physiological responses during the night. The consecutive 30-s epochs of PSG data collected from PSG sensors are then analyzed by experienced sleep physicians for achieving manual sleep stage classification. As we know, sleep staging involves categorizing sleep into different stages based on the patterns of EEG, EMG, EOG, and other physiological signals. The various physiological signals used in sleep research are broken down as follows:

- **EEG**: EEG is a technique designed to capture the electrical activity in the brain. This method entails the placement of electrodes on the scalp, which then recog-

nize and measure the brain's electrical signals. In the context of sleep studies, EEG is frequently employed to monitor and analyze brain wave patterns, facilitating the identification of distinct sleep stages, including the REM sleep stage and various NREM stages.

- **EMG**: EMG is a method designed to measure the electrical activity produced by muscles. Surface electrodes or fine wire electrodes are placed on the skin directly above the targeted muscles. In the context of sleep studies, EMG serves to monitor muscle tone, particularly in regions like the chin and leg muscles. By doing so, it aids in the differentiation of various sleep stages and facilitates the identification of conditions such as sleep disorders and movements occurring during sleep.

- **EOG**: EOG is a technique utilized to monitor the electrical activity resulting from eye movements. Electrodes are commonly positioned around the eyes to effectively monitor the motion of the eyeballs. EOG plays a crucial role in identifying the REM stage during sleep, as it is a distinct hallmark of REM sleep.

- **Other physiological signals (i.e., ECG)**: ECG records the electrical activity of the heart. Electrodes are placed on the chest and limbs to capture the heart's electrical signals. In sleep studies, ECG can provide information about changes in heart rate and rhythm during different sleep stages and can help identify sleep-related cardiac issues.

Among these PSG recordings, EEG stands out as a cost-effective and typically non-invasive method for monitoring and recording electrical brain activity during sleep. Additionally, EMGs and EOGs have been widely employed as crucial indicators for detecting the REM sleep stage [42]. As far as we know, the gold standard for sleep stage classification is overnight PSG. Therefore, correct sleep stage scoring is a key ingredient in this sleep analysis.

2.3.2 Scoring of Sleep Stages

According to the biological signals of overnight PSG recordings [43], human experts reach a manual scoring of sleep stages. At an early stage, the Rechtschaffen and Kales (R&K) criteria [44] for scoring of sleep stages is a widely acknowledged and standardized approach utilized to classify sleep stages, relying on PSG data. This system was formulated during the 1960s by renowned sleep researchers William C. Dement, Allan Rechtschaffen, and Anthony Kales. The R&K system involves the visual analysis and scoring of various physiological signals recorded during a sleep study. These signals include EEG, EOG, EMG, and other physiological signals. The trained sleep technicians or specialists use specific criteria to determine the sleep stage an individual is in at various points throughout the night. The R&K criteria divide sleep into six sleep stages, including the wakefulness (W) stage, the NREM sleep stage, and the REM sleep stage. In this context, the NREM category is further subdivided into four sleep stages: S_1, S_2, S_3, and S_4, respectively. The R&K criteria have been widely used for several decades and provide a standardized way to report sleep stages in research and clinical settings. However, the R&K criteria have several limitations, such as subjectivity in visual scoring and potential variability in interpretation between different scorers. Additionally, as our understanding of sleep physiology and technology advances, more detailed information on sleep stages becomes desirable. In order to address these limitations, the American Academy of Sleep Medicine (AASM) [45] periodically updates the sleep staging guidelines to incorporate new scientific knowledge and advancements in the field of sleep medicine based on the R&K criteria. These updates ensure that the guidelines remain accurate, relevant, and reflective of the latest understanding of sleep physiology and disorders. While the R&K system may not be as commonly used today, it remains an important historical reference in the field of sleep medicine. Researchers and clinicians who study the history of sleep staging and sleep research often refer to the R&K system to understand the evolution of sleep stage classification. In contrast, the AASM staging manual provides a more detailed and specific set of criteria for scoring sleep stages. In addition, the AASM manual also offers a more comprehen-

sive framework for sleep staging in clinical practice and research, which has become the contemporary standard for scoring sleep stages. According to the AASM manual, sleep experts use consecutive 30-s epochs of PSG data to classify five stages. These are wakefulness, rapid eye movement (also referred to as stage R), and three NREMs, N_1, N_2, and N_3. Based on the R&K criteria or the AASM manual, sleep stages are shown in Figure 2.4. As we know, different sleep periods are characterized by distinct physiological and neurological patterns. Therefore, Sleep experts often use various metrics and characteristics of different sleep stages to assess sleep quality and create a scoring of sleep stages.

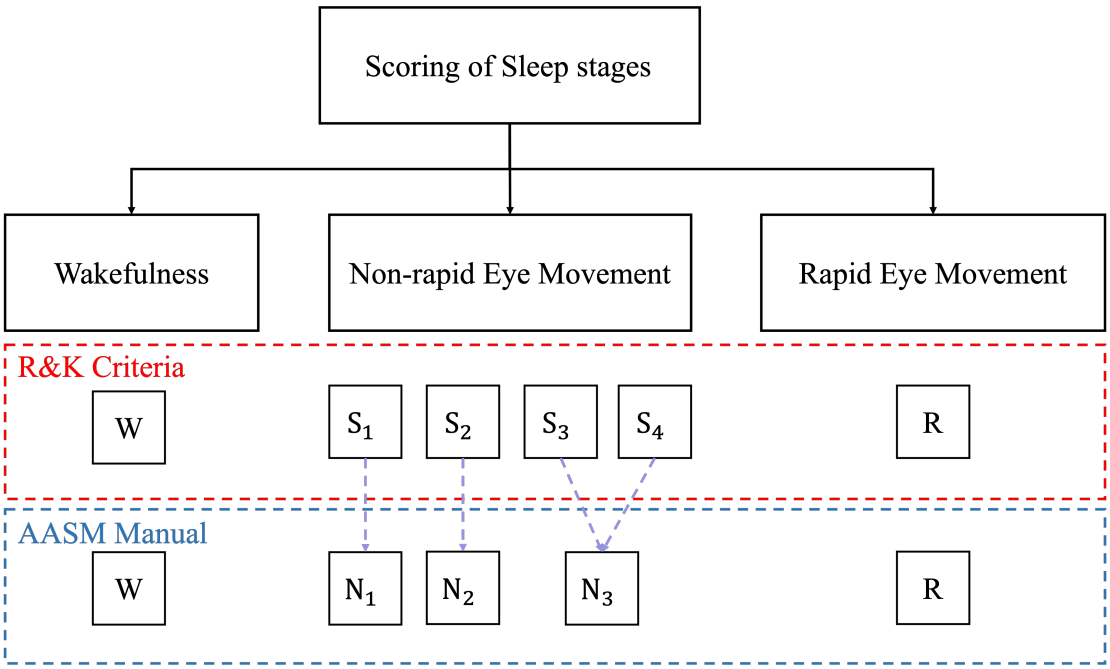


Figure 2.4: **Terminology used by R&K and AASM for sleep stage classification. In R&K criteria, the sleep stages are classified into W (wakefulness), S_1, S_2, S_3, S_4, and R (rapid eye movement). In the AASM manual, S_3 and S_4 stages are merged into a single stage N_3.**

2.3.3 Sleep Stages and Their Characteristics

Sleep is divided into five sleep stages, each with its own unique features and functions. As of our last knowledge update in September 2021, the AASM sleep staging manual includes the following five sleep stages:

1. Wakefulness: Similar to the R&K criteria, this stage corresponds to periods when

the individual is awake and alert. When people are awake and alert, brain activity is characterized by beta waves, which are fast and desynchronized. Muscle activity is present, and eye movements are rapid. During wakefulness, people are conscious and aware of their surroundings.

2. NREM sleep stage:

    - N_1 stage: N_1 denotes the transition stage between wakefulness and sleep. Similar to the S_1 stage of the R&K criteria. It's a very light sleep stage, and people can be easily awakened during this phase. It often lasts only a few minutes. During this stage, eye movements are slow, and muscle activity decreases.
    - N_2 stage: N_2 signifies the onset of true sleep, which is a deeper sleep than N_1 and is characterized by a decrease in heart rate and body temperature. Similar to the S_2 stage of the R&K criteria, this stage is characterized by a decrease in muscle activity and conscious awareness. Moreover, brain wave activity consists of short bursts of electrical activity. Sleep spindles and K-complexes, which are sudden spikes in brain wave activity, are common in this stage.
    - N_3 stage: N_3 is also known as slow-wave sleep (SWS) or deep sleep, which is the deepest stage of sleep. One of the significant changes introduced by the AASM manual in their updated sleep staging guidelines is the consolidation of the two deep sleep stages, S_3 and S_4 from the R&K criteria, into a single stage called N_3. It's characterized by slow brain waves (delta waves) and is considered the most restorative stage. Tissue repair and growth occur during this period, and it's often difficult to awaken someone from this stage.

3. REM sleep stage: Similar to the REM stage in the R&K criteria, it is characterized by rapid eye movements, increased brain activity, and vivid dreams. Physiologically, REM sleep resembles wakefulness. Brain activity produces rapid and

desynchronized beta waves. Eyes move rapidly in various directions, and this is when vivid dreaming occurs. Muscles are mostly paralyzed to prevent acting out dreams (atonia). REM sleep is crucial for emotional processing, memory consolidation, and cognitive functions.

Throughout the night, a typical sleep cycle consists of multiple cycles through these different stages, progressing from N_1 to N_2 to N_3 and then back to N_2 before entering REM sleep. Each cycle lasts around 90 minutes, with REM sleep becoming longer and more prominent in the latter cycles. This cycling between sleep stages is essential for overall sleep quality and restoration. Hence, Sleep experts use the characteristics of these different sleep periods, including the proportions of each stage, to assess sleep quality, diagnose sleep disorders, and make recommendations for improving sleep habits. However, clinical sleep scoring necessitates the meticulous visual examination of overnight PSG data by a skilled human expert in alignment with established criteria. Consequently, the manual classification of sleep stages demands substantial labor, consumes a significant amount of time, is intricate in nature, and carries the potential for errors [46]. Therefore, achieving a reliable and high-precision methodology for automatic sleep stage classification, utilizing bioelectrical signals, holds significant prominence within the domain of sleep research.

Chapter 3

Related Works

3.1 Deep Learning on Graph

With the increasing prominence of deep learning, the recent successes achieved through neural networks have boosted research in the domains of pattern recognition and data mining. Deep learning techniques present better performance than the state-of-the-art algorithms in many domains, including image classification [47–49], video processing [50–52], speech recognition [53–55], and natural language understanding [56–58]. Various end-to-end deep learning paradigms, like CNNs [59], RNNs [60], and autoencoders [61], have taken over the traditional approach of manual feature engineering. The success of deep learning in extracting latent representations from Euclidean data has proven that it can perform complex tasks with minimal human intervention [62, 63]. In particular, CNN can efficiently achieve image analysis by automatically learning and extracting features from images. Therein, this type of image data is specially processed by CNN and has a fixed spatial structure, which belongs to the two-dimensional grid data. However, not all significant real-world data is presented in the form of visual signals or two-dimensional or three-dimensional information. These kinds of data are collectively known as graph data. Graphs are pervasive in the real world, manifesting in various domains and scenarios. Examples of graph data include social networks [64], knowledge graphs [65], protein-protein interaction networks [66],

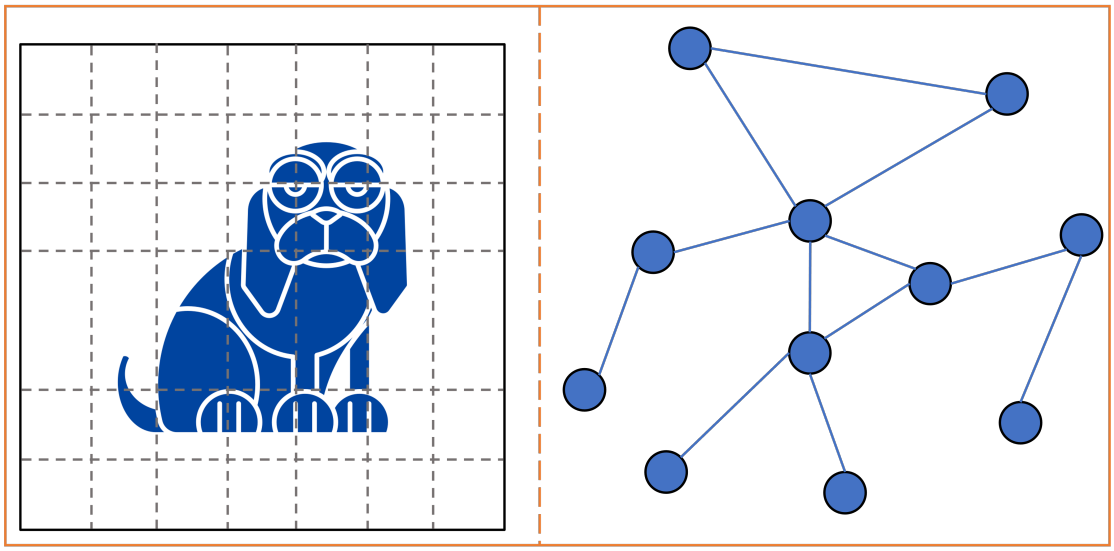


Figure 3.1: **Left: image in Euclidean space. Right: graph in non-Euclidean space.**

and so on. In computer science and mathematics, a graph is a data structure that consists of two sets, i.e., node (or vertex) set and edge set. The nodes represent entities in a graph, whereas edges represent relationships between those entities. Graphs are used to model relationships between various entities. As shown in Figure 3.1, localized convolutional filters and pooling operators are not easy to define. It hiders the transformation of CNN from the Euclidean domain to the non-Euclidean domain.

Traditional CNN-based methods are limited in their ability to effectively process graph data [67]. In order to effectively process and analyze such graph data, graph learning [68] has been introduced to extract intricate relationships from graph-structured data by leveraging essential and relevant relations among vertices within the graph. For example, the detecting information cascades are used to track the spread trajectory of rumors in social networks. Some researchers [69] analyze the co-occurrence phenomenon with different timestamps to achieve human mobility pattern prediction in traffic networks. Thus, a massive amount of the existing studies show that extracting complex patterns by exploiting deep learning from graph data has been found very useful in various fields, such as pattern recognition and image processing. As deep learning techniques can encode and represent graph data into vectors, how to utilize deep learning techniques to extract patterns from complex graphs has attracted considerable attention.

Deep learning on graphs can be divided into three categories: semi-supervised, un-

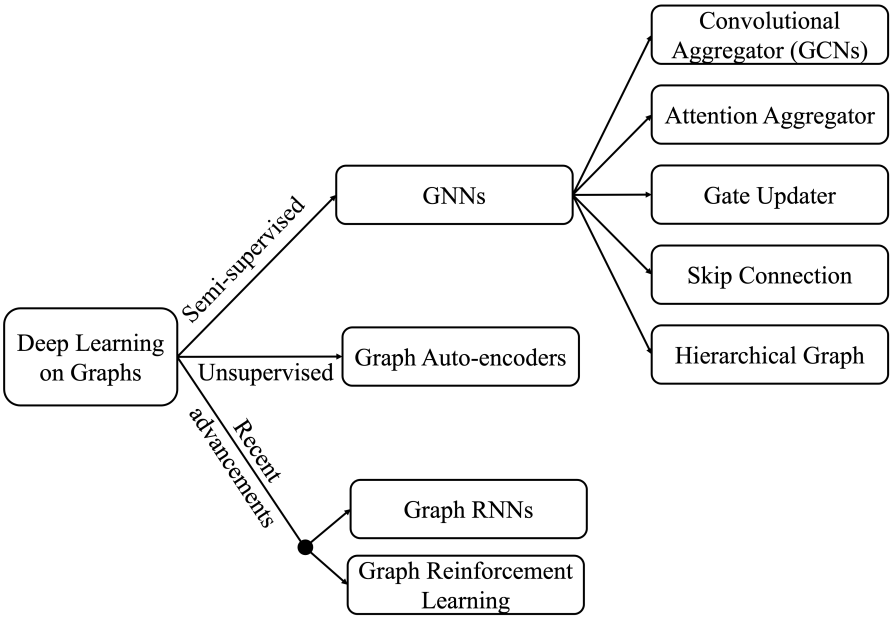


Figure 3.2: **The categorization of deep learning-based methods on graphs.**

supervised, and recent advancements. GNNs [70] is a key component of deep learning techniques specifically recognized as a powerful technique for graph analysis, which falls under the umbrella of graph learning in the context of deep learning. GNN is also an effective framework for processing data that can be represented as graphs [71]. In the work proposed by Wu et al. [72], it has been proved that some state-of-the-art GNN models are successfully applied in various fields. GNNs are classified into five groups: convolutional aggregator [3], attention aggregator [73], gate updater [74,75], skip connection [76], and hierarchical graph [77], with GCNs being only a small subset in this broader context. GCN falls under the category of convolutional aggregators, as shown in Figure 3.2. In addition, GCNs are proposed to extend CNNs to graphs, allowing for the effective modeling of relationships and dependencies within graph data. Therefore, GCNs and subsequent variants have become the mainstream models used for learning graph representations and achieving excellent performance.

3.2 Graph Convolutional Network

GCNs [2] are a type of multilayer neural network architecture designed to process and analyze graph-structured data. GCNs function directly on a graph and generate

embedding vectors of nodes according to the properties of their neighborhoods. The existing graph convolutional network model can be divided into two taxonomies, as shown in Figure 3.3. First taxonomy is contingent on the type of convolutions, GCNs can be categorized into two main models: spectral-based models and spatial-based models. Another classification can be based on the practical applications in which GCNs are employed. From this, it is evident that GCNS have applied in a wide range of tasks and applications. In the GCN framework, each part can be introduced as follows:

- **Input Data**: The graph is represented by an adjacency matrix, which captures the relationships between nodes, and a feature matrix, which contains feature vectors for each node.

- **Convolutional Operation**: GCNs use a localized convolutional operation similar to CNNs. However, in GCNs, this operation is adapted to work on the graph structure. It involves aggregating and transforming information from neighboring nodes.

- **Message Passing**: Nodes in the graph gather and exchange information with their neighbors. This information exchange is called message passing. Each node aggregates information from its neighbors' features and updates its own feature representation.

- **Layer Stacking**: GCNs are typically composed of multiple layers. Each layer refines the node representations by incorporating information from increasingly distant nodes in the graph.

- **Activation Function**: After each layer, an activation function (often using ReLU) is applied to the updated node features.

- **Output**: Depending on the task, the final layer's node features can be used for various purposes, such as node classification, link prediction, or graph classification.

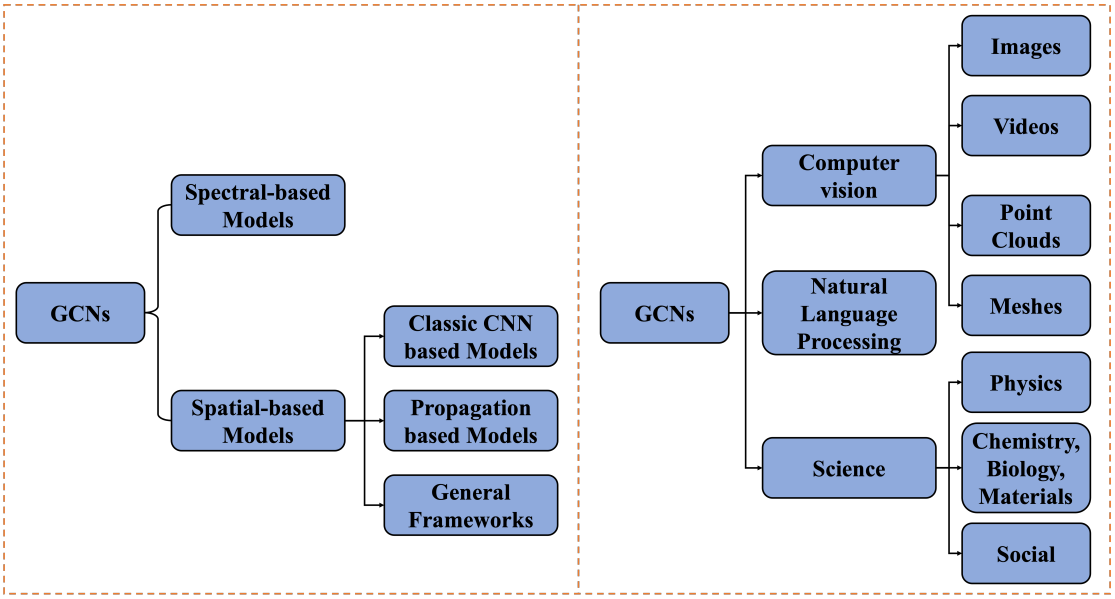


Figure 3.3: **An overview of graph convolutional networks. Left: based on the types of convolutions. Right: based on the application domains.**

We briefly review the development of GCNs. In 2013, it was widely recognized that Bruna et al. [78] were among the first to introduce spectral-based GCN. Defferrard et al. [79] uses K-polynomial filters to improve the computational complexity. In addition, the Chebyshev polynomial [80] is employed to future reduce the computational complexity. Thereafter, a simplified GCN model [2] is proposed for semi-supervised classification and to achieve better predictive performance on a number of datasets. However, the spectral graph convolution relies on the specific eigenfunctions of the Laplacian matrix. It remains challenging to transfer the spectral-based graph convolutional network models learned on one graph to another graph with different eigenfunctions. Hence,spatial-based models might be favored. For example, Gao et al. [81] propose the learnable GCN model to handle more irregular structures. In [82], the topology adaptive graph convolutional network is adaptive to the graph topology as the filter scans the graph.

On the other hand, GCNs can be also categorized according to their application domains. Yao et al. [83] present a novel architecture combining GCN and long short-term memory (LSTM) for image captioning. In [84], authors use graph convolution to process input graphs for generating images from scene graphs. Moreover, in our

work [85], we develop a lightweight yet efficient neural network built upon GCN for action recognition based on skeleton data extracted from action clips. Another skeleton-based method is [86], where a generalized GCN is proposed for better recognition in skeleton-based action recognition exploiting spatial and temporal dependencies among human joints. To accomplish point cloud classification, Wang et al. [87] have introduced a local spectral-based graph convolution method designed to achieve a specific objective. The [88] presents a novel graph-convolutional architecture named FeaStNet to find correspondences between collections of 3D shapes on meshes. Generally speaking, text classification is an important and classical problem in natural language processing. To address this task, Many GCN-based models have been proposed, to name a few, the Text GCN [89] is proposed for text classification and its performance surpasses that of many methods. Furthermore, Lin et al. [90] design a text classification model named BertGCN utilizing the advantages of GCN and bidirectional encoder representations from transformers (Bidirectional Encoder Representations from Transformers (BERT)) to achieve state-of-the-art performances on a wide range of text classification datasets. GCNs have been applied in the relation extraction between words [91] and event extraction [92]. GCNs have found applications in various scientific domains, such as the physical dynamics [93], chemical stability prediction of a compound [94], and protein interface prediction [66, 95, 96]. In particular, the GCNs are applied directly to classify sleep stages. In the context of EEG data, as brain regions exist in a non-Euclidean space, a graph is the most suitable data structure for representing brain connections. For the EOG data, we create a graph representation that captures relationships between EOG channels. Therein, each electrode channel in the EOG data can be represented as a node in the graph. The edges represent relationships between nodes, namely functional connectivity. As a consequence, our proposed second work [97], 4s-SleepGCN, employs EEG and EOG signals for sleep stage classification based on GCN. In summary, GCNs have gained popularity for their ability to capture both local and global information from graph data, making them useful for a wide range of tasks in fields like social network analysis, sleep stage classification, recommendation systems, molecular

chemistry, and more.

In spite of the fact that GCNs are a powerful tool for processing graph-structured data. However, spatiotemporal data often involves complex interactions between spatial and temporal dimensions, which may not be adequately captured by traditional GCNs. Hence, the ST-GCN, as one of the most advanced extensions of GCN-based models, can introduce spatiotemporal convolutional layers to incorporate both spatial and temporal information by leveraging the neighborhood relationships between nodes and the sequential dependencies across time steps.

3.3 Spatiotemporal Graph Convolutional Network

In recent years, there has been a growing interest in leveraging deep learning techniques to effectively model spatiotemporal data. ST-GCN [98] represents a cutting-edge class of deep learning models designed to process spatio-temporal data in complex systems. ST-GCN leverages graph convolutional operations to capture both spatial and temporal dependencies in data sequences. It allows the modeling of dynamic interactions among entities in a dataset, making it suitable for a wide range of applications. For instance, in the field of computer vision, action recognition, and human pose estimation. This section reviews the key contributions and advancements in the field of ST-GCN, providing a comprehensive understanding of their capabilities and applications.

3.3.1 Fundamentals of ST-GCN

ST-GCN fundamentally expands upon the principles of GCNs to accommodate spatiotemporal data. ST-GCN is built on the foundation of graph theory and deep learning, enabling the modeling of data structured as graphs. To put it differently, the core of ST-GCN lies in the fundamental idea of representing data as a graph [99]. A graph consists of nodes (vertices) and edges (connections) that define the relationships between nodes. This representation is particularly suitable for data with intrinsic spatial connections, such as urban road networks [100] or social networks [101]. In ST-GCN,

nodes represent individual entities within the data, which may be locations, sensors, or even individuals in a social network. In addition, edges represent relationships or interactions between nodes. These relationships can be directed or undirected. Therefore, the spatial dependencies can be captured by considering neighboring nodes in the graph structure. Moreover, since each graph corresponds to a different time step in the sequence, ST-GCN operates on the temporal sequence of the graph for capturing temporal dependencies.

ST-GCN is a combination of the temporal convolutional network (TCN) [102] and GCN [103]. TCN conducts convolutional operations on data in the temporal dimension, while GCN performs convolutional operations on data in the spatial dimension. Spatial convolutional layers are applied to each graph in the sequence. These layers compute feature representations by considering the relationships between nodes in the same graph, capturing spatial dependencies within each frame. In spatial convolution, the model performs convolutions on each graph (representing one frame or snapshot in the temporal sequence). It considers the local neighborhood of each node in the graph, along with their features, to compute new feature representations. This process captures spatial dependencies within a single frame. The spatial convolutional operation can be represented as:

$$Y_s = \varpi \left(\sum_{i=1}^{P} \sum_{j \in \mathbb{P}(i)} W_s \cdot V_i \cdot V_j \right) \tag{3.1}$$

where Y_s denote the output feature map. P is the number of nodes in the graph. $\mathbb{P}(i)$ represents the neighborhood of the node. W_s are learnable spatial convolutional filters. V_i and V_j denote node features. The activation function ϖ commonly uses ReLU.

TCN specializes in modeling sequential data, particularly time-series data. It is designed to capture long-range temporal dependencies efficiently. TCN uses a series of 1D causal convolutions [104], which are dilated to capture different temporal resolutions. Dilated convolutions [105] allow TCN to model temporal dependencies at various scales and effectively handle long-range dependencies. Temporal convolutional

layers are applied across the sequence of graphs. These layers capture how features change over time, modeling temporal dependencies, which can be represented as:

$$Y_t = \varpi \left(W_t * X_t \right) \tag{3.2}$$

Therein, Y_t is the output feature map. W_t and X_t denote learnable temporal convolutional filters and the sequence of spatial feature maps over time, respectively. Therefore, by combining spatial and temporal convolutions, ST-GCN effectively captures both spatial and temporal dependencies in spatio-temporal data, making it well-suited for tasks such as action recognition and human pose estimation in videos.

3.3.2 Applications of ST-GCN

ST-GCN has a wide range of applications across various domains due to its ability to capture both spatial and temporal dependencies in data. Some key applications of ST-GCN are presented as follows:

1. **Action Recognition**

   In the skeleton-based human action recognition domain, the methods based on ST-GCN have had great success recently. Yan et al. [98] propose an ST-GCN-based method to model the skeleton data and this method improves the accuracy of action recognition to a new level. In [106], an improved ST-GCN model is proposed to well capture the intrinsic high-order correlations among skeleton joints. Moreover, Li et al. introduce an actional-structural graph convolutional network (AS-GCN) [107], which contains actional-structural graph convolution and temporal convolution, to capture richer dependencies. Shi et al. [108] take advantage of the relationship between joints and bones for action recognition. In addition, more variants of ST-GCN are proposed [85, 109–111], which typically introduce incremental modules to enhance the expressiveness and network capacity.

2. **Human Pose Estimation**

ST-GCN can estimate the 2D or 3D poses of human bodies over time, enabling applications in animation, sports coaching, and healthcare for monitoring patient movements. For example, a method [112] based on an improved ST-GCN model is proposed for human pose estimation. In [113], the authors use hand pose sequences as input and estimate 3D hand joint locations using ST-GCN. Furthermore, Sofianos et al. [114] present a novel space-time-separable graph convolutional network (STS-GCN) to better learn the fully trainable joint and time interactions for pose forecasting. Above that, Liu et al. [115] use temporal convolutional and graph attention blocks to capture varying spatiotemporal sequences for achieving real-time 3D human pose estimation in video. Overall, ST-GCN provides an effective framework for human pose estimation by considering both spatial and temporal information, making it robust in scenarios where poses change dynamically over time.

3. **Healthcare**

In healthcare, ST-GCN can be used for gait analysis, assessing the quality of movement in rehabilitation exercises, sleep stage classification, and monitoring patient activity for fall detection or eldercare. For instance, Keskes et al. [116] develop an effective fall detection system that offers good results using ST-GCN. In [117], The researchers employ the ST-GCN model to predict the probability of a person falling or not falling. Moreover, the ST-GCN-based approach introduced by Lu et al. [118] is designed to learn and capture the left ventricular (LV) motion patterns, illustrating the adaptability of ST-GCN in medical imaging applications. In the domain of sleep stage classification, our initial work [119] employs a combination of dynamic and static ST-GCN with inter-temporal attention blocks to achieve state-of-the-art performance. Zhao et al. [120] utilize GCN and TCN to extract spatial features and transition rules between sleep stages. In the Internet of Medical Things (IoMT), ST-GCN can monitor patient activity and behavior, ensuring patients are following prescribed routines and identifying irregularities or distress signals [121]. Therefore, in healthcare, ST-GCN's ability to analyze

spatio-temporal data can contribute to early diagnosis, effective rehabilitation, and improved patient care.

In summary, the versatility of ST-GCN highlights its capacity to capture intricate spatio-temporal patterns and dependencies. This makes it a valuable tool for comprehending and analyzing complex data across various domains and applications. Simultaneously, an increasing number of individuals are placing their attention on the performance of these networks. In order to enhance the network's performance in various tasks, an attention mechanism is introduced in the neural network.

3.4 Attention Mechanism

Attention is a complex cognitive function that plays an indispensable role in human behavior and perception [122,123]. Humans selectively focus their attention on specific information when and where it is required. Meanwhile, they ignore some perceivable information. This is a mechanism for humans to expeditiously extract valuable data from vast information within limited cognitive resources. The attention mechanism greatly improves the efficiency and accuracy of perceptual information processing.

The attention mechanism [124] is a technique for diverting attention to the most important regions while filtering out extraneous or inconsequential areas. In deep learning, the attention mechanism is a key component, particularly in the field of natural language processing (NLP) and computer vision, that allows these models to focus on specific parts of the input data when making predictions or decisions. In the human visual system, attention [125] is used as an aid to efficiently analyze and understand complex scenarios. Inspired by the idea of attention, the attention mechanism is introduced to improve performance in the field of computer vision. Within a visual processing system, an attention mechanism can be considered as a dynamic selection procedure, wherein features are adaptively assigned weights based on the significance of the input. The actual starting point of the attention mechanism is developed for sequence-to-sequence tasks and quickly becomes popular for a variety of visual tasks, such as image classifica-

tion [126], object detection [127], semantic segmentation [128], face recognition [129], super resolution [130], sleep stage classification [119], medical image processing [131], and multi-modal task [132].

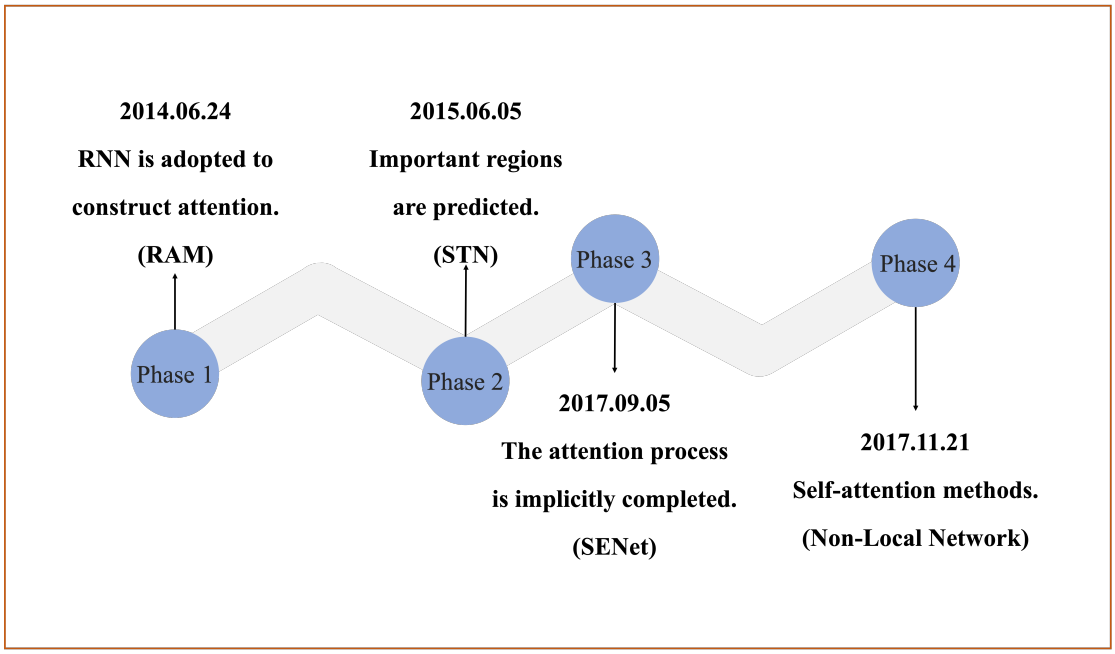


Figure 3.4: **Brief summary of key developments in attention in computer vision, which have loosely occurred in four phases. A representative method in each phase is RAM [132], STN [132], SENet [124], and Non-Local Network [133], respectively.**

The history of attention-based deep learning models in computer vision can be coarsely divided into four phases, as shown in Figure 3.4. The initial phase commences with the introduction of the recurrent attention model (RAM) [133], a groundbreaking endeavor that integrated deep neural networks with attention mechanisms. At the start of the second phase, the spatial transformer network (STN) [134] is proposed to select important regions in the input using the spatial transformer. The third stage was inaugurated by the squeeze-and-excitation network (SENet) [126], which introduced an innovative channel-attention network capable of implicitly and adaptively forecasting potential key features. The final phase marks the advent of the self-attention era, which can be initially proposed in [135] and swiftly brings about substantial advancements in the NLP. The attention-based models show the huge potential in various tasks. It is evident that attention-based models hold the promise to supplant convolutional neural networks and emerge as a more robust and versatile architectural paradigm in the realm

of computer vision. There are two main types commonly used in convolutional neural networks: spatial attention and channel attention.

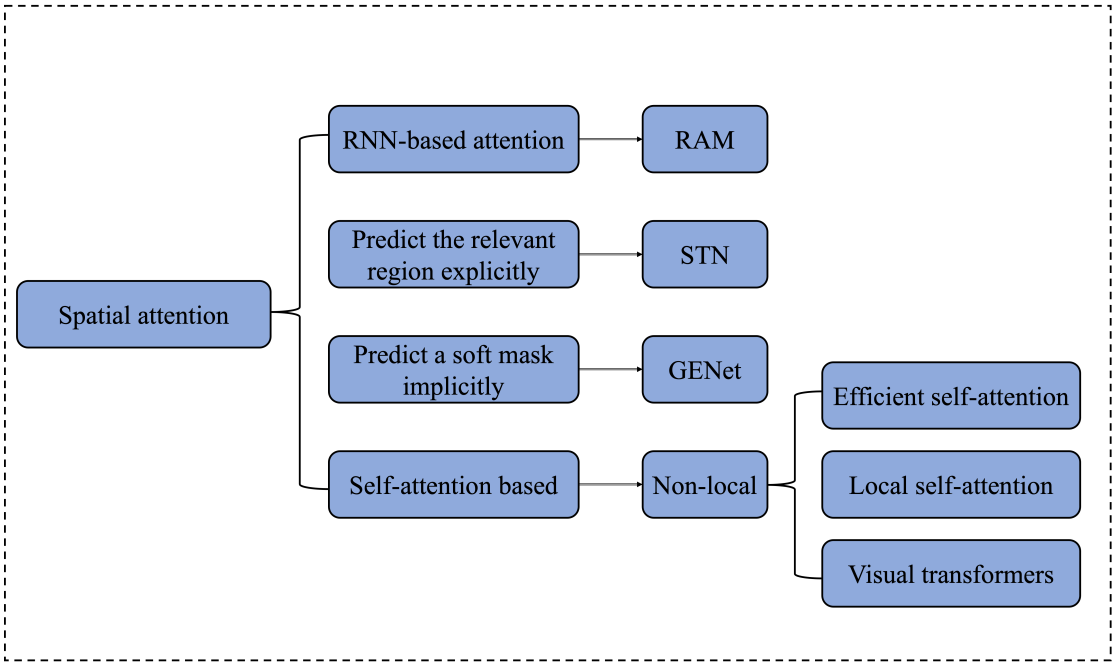


Figure 3.5: **Developmental context of spatial attention methods.**

1. **Spatial Attention**

   Spatial attention can be seen as an adaptive spatial region selection mechanism: where to pay attention. As illustrated in Figure 3.5, RAM [133], STN [134], gather-excite network (GENet) [136], and Non-Local [135] serve as exemplary instances of various spatial attention methodologies. It is well known that CNNs have huge computational costs, especially for large inputs. In order to concentrate limited computing resources on important regions, RAM incorporates RNNs [137] and reinforcement learning [138] to instruct the network on where to focus its attention. RAM represents a pioneering effort in employing RNNs for visual attention and paves the way for subsequent RNN-based techniques, as evidenced by the subsequent works [139, 140]. Moreover, the attribute of translation equivariance renders CNNs well-suited for handling image data. Nevertheless, CNNs do not possess rotational invariance, scaling invariance, and warping invariance. To achieve these attributes while making CNNs focus on important regions, STN

is proposed to enable the network to focus on discriminative regions by using an explicit procedure and grant deep neural networks the ability to achieve transformation invariance. For GENet, the long-range spatial contextual information can be extracted through the incorporation of a recalibration function in the spatial domain. The architecture of GENet integrates both part-gathering and excitation operations. The proposed gather-excite module captures important features while simultaneously suppressing less relevant information. Self-attention has been successfully applied in the field of NLP and is employed as a spatial attention mechanism to capture global information features. CNNs have a limited receptive field because the convolutional operations are localized. To address this problem, the self-attention is introduced in computer vision [135]. Nonetheless, the self-attention mechanism is afflicted with various shortcomings, with its quadratic computational complexity being a notable constraint on its practicality. In response to these challenges, several variants (i.e., CCNet [141], A^2Net [142], and SAN [143]) have been introduced to mitigate these issues. In addition, the vision transformer architecture [144] is used in image processing, which is capable of producing results equal to those of modern convolutional neural networks.

In summary, spatial attention mechanisms in deep learning enable models to focus on specific regions or parts of input data, improving their ability to process complex information and make more accurate predictions. These mechanisms have become integral components of various state-of-the-art deep learning architectures and have been crucial in advancing the performance of models in a wide range of tasks.

2. **Channel Attention**

In deep neural networks, different channels in different feature maps usually represent different objects [145]. The weight of each channel can be adaptively recalibrated through channel attention, akin to a process of selecting objects or features, consequently determining where to allocate attention. Hence, SENet [126] represents a pioneering advancement in the realm of channel attention. Therein,

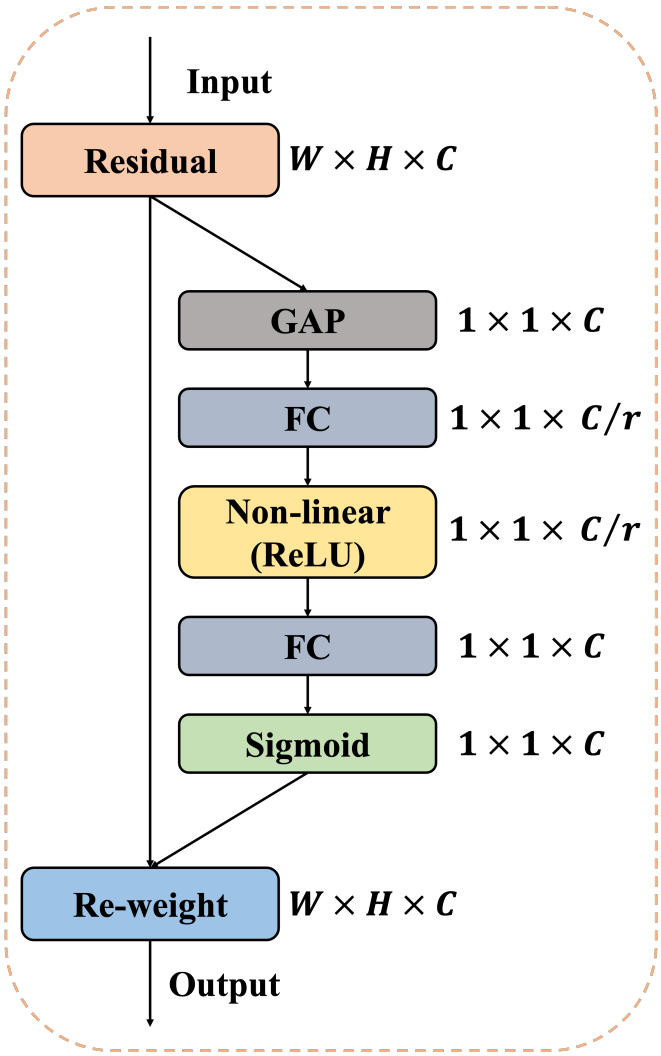


Figure 3.6: **The Schema of the SE Block. GAP and FC denote global average pooling and fully connected layer, respectively.**

a SE block in SENet serves the crucial functions of collecting global information, capturing inter-channel relationships, and enhancing the network's representation capabilities. SE blocks (see Figure 3.6) contain a squeeze module and an excitation module. The squeeze module aggregates global spatial information through global average pooling. Meanwhile, the excitation module captures inter-channel relationships and generates an attention vector through the utilization of fully connected (FC) layers and non-linear functions. Subsequently, each channel within the input feature is scaled by multiplying it by the corresponding element from the attention vector, effectively building interdependencies between channels. While the SE block is widely used, it solely re-evaluates the importance of each channel

by modeling channel relationships, neglecting positional information. However, positional information is critical for generating spatially selective attention maps. To address this problem, the coordinate attention (CA) block is introduced as a novel attention block based on the SE block that takes into account not only the relationship between channels but also the positional information in the feature space.

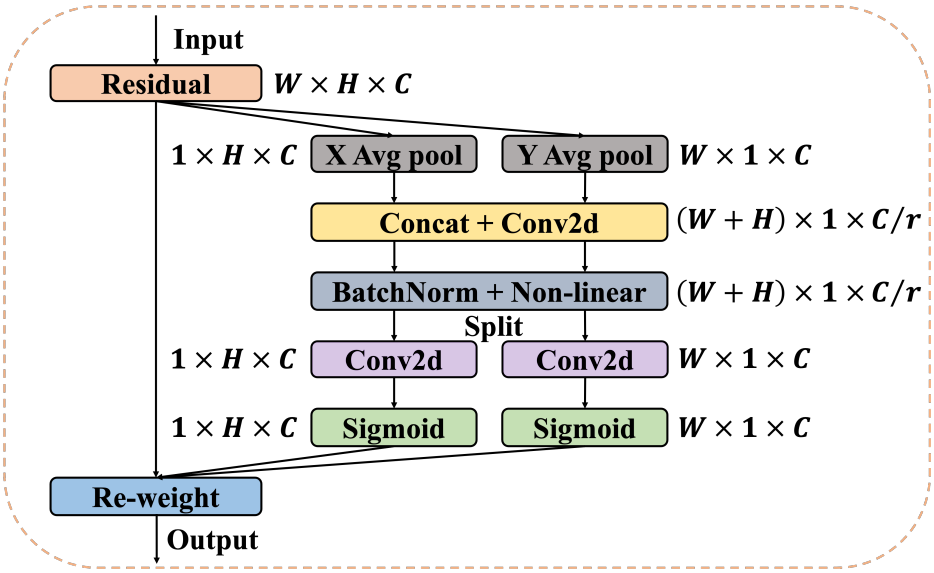


Figure 3.7: **The Schematic of the CA Block. X Avg Pool and Y Avg Pool refer to 1D horizontal global pooling and 1D vertical global pooling, respectively.**

Coordinated attention encodes channel relationships and long-range dependencies with precise positional information. The specific operations are divided into coordinate information embedding and CA generation, as shown in Figure 3.7. Coordinate information embedding aggregates features along two spatial directions separately, yielding a pair of directionally aware feature maps. Additionally, it enables attention modules to capture long-range dependencies along one spatial direction while preserving precise positional information along the other spatial direction. This assists the network in more accurately localizing the regions of interest. In the CA generation, not only does it make full use of the captured positional information, accurately capturing the regions of interest, but it also effectively extracts inter-channel relationships. Therefore, CA blocks can be incorporated into various neural network architectures to improve performance

by effectively capturing global context information. This is especially true for classification tasks such as sleep stage classification, where understanding the relationships between different temporal coordinates or channels is critical.

Attention mechanisms are powerful tools for modeling relationships within a single modality or data source, as they enable models to focus on specific parts or elements within that modality, which can help capture important patterns or dependencies. However, when dealing with multiple modalities or data sources, such as text, images, audio, or other types of data, it becomes necessary to model the relationships and interactions between these modalities. To overcome this limitation, multimodal fusion is proposed, enabling the integration of data from diverse sources to deduce valuable information that may not be attainable through the use of a single source.

3.5 Multimodal Fusion

Multimodal [146]refers to the integration or interaction of multiple sensory modalities or data sources. In the context of data analysis, machine learning, and artificial intelligence, multimodal data or systems involve the combination of information from different sources or modes. These modalities can include various types of sensory input or data types, such as text, images, video, sensor data, environmental data, biometric data, and geospatial data. Multimodal data analysis involves the processing, integration, and interpretation of information from these different modalities to gain a more comprehensive understanding of a given task or problem. For example, in sleep stage classification, a multimodal system [97] might use data from different PSG signals to accurately classify sleep stages. Therefore, multimodal fusion can enhance the robustness, accuracy, and richness of information processing in various applications, allowing systems to better mimic human perception and understanding.

Multimodal fusion [147] refers to the process of integrating information or data from multiple sensory modalities or sources to create a unified representation or understanding of a given phenomenon, event, or problem. The goal of multimodal fusion is

to combine the strengths of different modalities to enhance overall system performance and enable more comprehensive analysis. Multimodal fusion can occur at different stages of data processing, leading to three common approaches: early fusion, intermediate fusion, and late fusion [148], as shown in Figure 3.8.

3.5.1 Fusion Structure

3.5.1.1 Early Fusion

Early fusion in multimodal fusion is a technique where information from different modalities or data sources is combined at the input level before being processed by a single model. This approach creates a joint representation of the data from multiple sources, allowing the model to learn interactions and dependencies between modalities from the very beginning of the processing pipeline. Poria et al. [149] propose an early fusion method that involves concatenation of multimodal features. However, the early fusion of multimodal data may not fully harness the complementarity of the modalities and could result in exceedingly large input vectors that might include redundant information. To address this problem, autoencoders [150] in deep learning are used for dimensionality reduction and feature learning. Moreover, in [151], the authors use some convolutional, training, and pooling fusion methods to solve the other challenge in early fusion, namely time synchrony between different data sources.

3.5.1.2 Late Fusion

Late fusion, also known as late integration or late combination, is a multimodal fusion technique where the information from different modalities or data sources is processed independently and then combined at a later stage. During the fusion stage, a variety of techniques are employed to fuse the outputs from the modality-specific models, ultimately producing a final multimodal prediction. For example, weighted summation [152] is applied to assign weights to the output of each modality and compute a weighted sum. These weights can either be learned from the data through training or set manually based on prior domain knowledge. The second common technique is con-

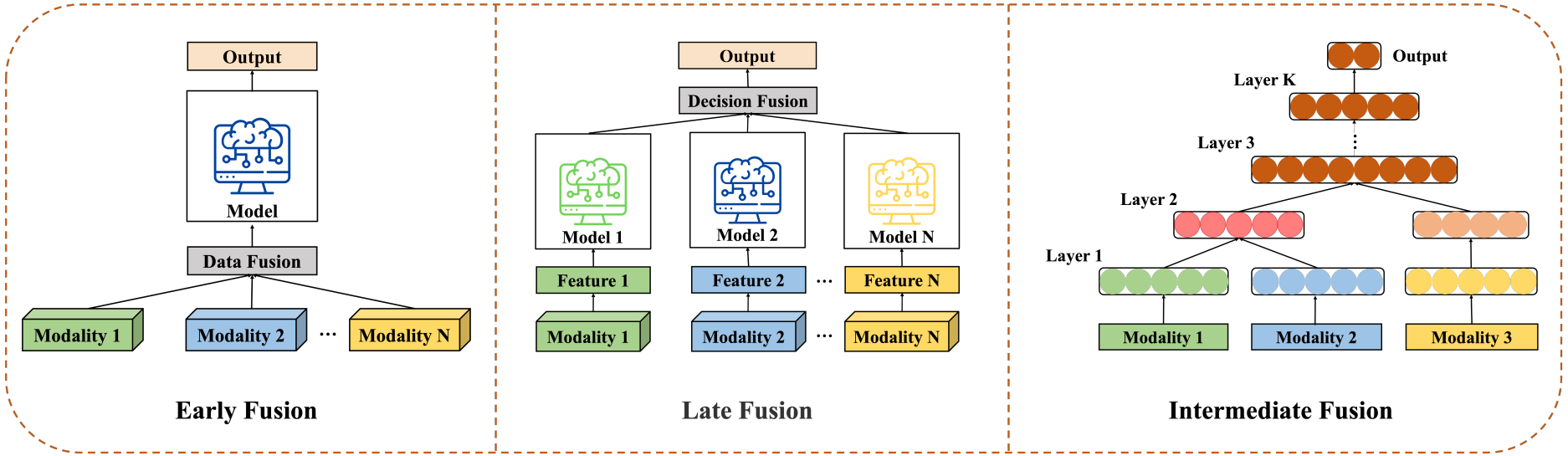


Figure 3.8: **An illustration of various fusion models for multimodal learning.**

catenation [153] which concatenates the outputs from different modalities into a single feature vector and then feeds it into a fusion layer or model. Furthermore, the predictions from modalities are combined using voting mechanisms (e.g., majority voting) or consensus algorithms [147]. Recently, attention mechanisms [140, 154] have been introduced to dynamically weigh the contributions of different modalities based on their relevance to the task. When input modalities exhibit substantial uncorrelation, considerable disparity in dimensionality, or distinct sampling rates, Implementing a late-fusion approach in multimodal learning problems is often more favored. In contrast, an intermediate fusion approach provides greater flexibility.

3.5.1.3 Intermediate Fusion

Intermediate fusion is a technique in multimodal deep learning where data from different modalities are first transformed into high-dimensional feature representations and then fused at an intermediate layer within a neural network model. The Inputs are scaled, shifted, and skewed by alternating linear and nonlinear operations at each layer to generate a new representation of the original data. In multimodal fusion, when raw data from various modalities are converted into high-dimensional feature representations, these feature representations are fused into a single hidden layer, allowing the model to learn a joint multimodal representation. During this processing, 2D convolution, 3D convolution, or FC layer is used for learning high-dimensional feature representations. After the individual modalities have been processed, the representations from these modalities are combined using a fusion layer, also known as a shared representation layer. This fusion layer is responsible for integrating the modalities into a joint or shared representation space. Finally, the fusion layer produces a unified, joint multimodal representation that captures the interactions and relationships between the modalities. The study [155] shows that intermediate fusion can be especially useful when there are complex interactions between data sources and additional processing is needed to align the data.

In summary, the choice of fusion approach depends on the specific application, the

nature of the data sources, and the desired trade-offs between computational complexity, information integration, and flexibility. Different fusion methods may be more suitable for different tasks and scenarios.

3.5.2 Applications of Multimodal Fusion

Multimodal fusion has a wide range of applications across various domains where information from multiple sources or modalities needs to be combined to make more informed decisions or improve the overall understanding of a complex phenomenon. Deep multimodal learning started to garner significant attention within the research community after the pioneering work of [156]. The initial endeavor in deep multimodal fusion primarily focused on only two modalities: images and text. As time has progressed, an increasing number of researchers have directed their efforts toward leveraging multimodal data for human activity recognition (HAR). Multimodal fusion in HAR combines multimedia data such as audio, video, depth, and skeletal motion information to create a holistic view of human activities. Multimodal deep learning approaches have been used to solve a wide array of problems related to human activity, such as skeleton-based action recognition [157], speech and gesture recognition [158], facial emotion recognition [159], and face recognition [160].

In addition, with the undeniable success of deep learning in medical applications, the multimodal fusion of information from different modalities to improve medical diagnosis, treatment, and health management has attracted great interest in medicine. Combining data from different sensors, imaging modalities, patient records, and other sources can lead to a more comprehensive and accurate understanding of a patient's condition. In disease diagnosis, multimodal fusion combines data from various medical tests and diagnostic tools, such as blood tests, medical imaging (e.g., magnetic resonance imaging (MRI), computerized tomography (CT) scans, X-rays), and patient history. For example, combining information from multiple imaging modalities (MRI and positron emission tomography (PET) scans) can improve the accuracy of tumor detection and staging [161]. Moreover, the structured data from EHRs are integrated with

unstructured clinical notes and reports to improve patient risk prediction and outcome analysis [162]. Multimodal fusion aids in extracting valuable insights from heterogeneous healthcare data. Also, multimodal fusion combines NLP techniques with medical imaging or genomic data [163] to enhance diagnostic accuracy and support clinical decision-making, thereby extracting meaningful information from clinical narratives, radiology reports, and medical literature. In sleep stage classification, multimodal fusion is used for PSG analysis, namely, fusion of PSG signals to accurately classify sleep stages and detect sleep disorders, e.g., our proposed 4s-SleepGCN [97]. As the medical community increasingly embraces diagnosis aided by artificial intelligence, we can anticipate significant advances in the field through multimodal fusion techniques.

To this point, our investigation highlights the related techniques and methodologies in our proposed work. These techniques serve as our inspiration and benchmark for innovation. Furthermore, we draw insights from related techniques that can enhance the effectiveness of our approach. We explore how techniques can be integrated to create a more holistic and innovative solution.

Chapter 4

An Attention-guided Spatiotemporal Graph Convolutional Network for Sleep Stage Classification

In this chapter, we propose a single-channel EEG-based approach to achieve sleep stage classification. Sleep stage classification has been widely used as an approach in sleep diagnoses at sleep clinics. GNN-based methods have been extensively applied for automatic sleep stage classifications with significant results. However, the existing GNN-based methods rely on a static adjacency matrix to capture the features of the different EEG channels, which cannot grasp the information of each electrode. Meanwhile, these methods ignore the importance of spatiotemporal relations in classifying sleep stages. In this work, we propose a combination of dynamic and static ST-GCN with inter-temporal attention blocks to overcome two shortcomings. The proposed method consists of a GCN with a CNN that takes into account the intra-frame dependency of each electrode in the brain region to extract spatial and temporal features separately. In addition, the attention block was used to capture the long-range dependencies between the different electrodes in the brain region, which helps the model classify the dynamics of each sleep stage more accurately. In our experiments, we used the Sleep-EDF-39 and the subgroup III of the ISRUC-SLEEP dataset to compare with

the most current methods. The results show that our method performs better in accuracy from 4.6% to 5.3%, in kappa from 0.06 to 0.07, and in macro-F score from 4.9% to 5.7%. The proposed method has the potential to be an effective tool for improving sleep disorders.

4.1 Introduction

In recent years, the intersection of neural networks and biomedical engineering has seen groundbreaking advances, paving the way for innovative applications in numerous health-related fields. One of these promising applications is in the realm of sleep research, particularly in sleep stage classification. Sleep is an indispensable physiological phenomenon for human beings that serves to prevent physical and mental illness and improve mood [164]. Accurate classification of these stages is essential for diagnosing and understanding sleep disorders and improving overall sleep quality.

The traditional approach to classifying sleep stages relies heavily on the EEG. The EEG is an inexpensive and generally non-invasive test for monitoring and recording electrical activity during sleep. In addition, EMGs and EOGs have been used as two important switches for detecting the REM sleep stage [165]. Up to now, the conventional visual scoring method is still the most acceptable approach, namely that human experts need to combine other biological signals (such as EEG, EOG, and EMG) to achieve manual sleep stage classification [166]. Manual sleep stage classification, an integral component of sleep analysis, is considered a tedious task [167]. Although qualitative sleep scoring is indispensable, it is beset by limitations, chiefly the variability in interpretations attributable to the differing experiences of experts. This can lead to inconsistencies in scoring outcomes, undermining the reliability of the results. Moreover, manual visual inspection of an entire night's EEG data is an extremely time-consuming task. Considering these challenges, automatic sleep stage classification with rapid and high accuracy based on EEG signals is of great research interest.

Looking back on the past decades, various methods in the relevant studies on sleep

stage classification have been proposed. According to study [164], sleep stage research has far-reaching implications for biomedical practice. In the early days, researchers used hand-engineered feature-based methods to extract features in the time and frequency domains for sleep stage analysis. For example, Tsinalis et al. [168] made the precision of sleep stage classification up to 78.9% via the extracted features in the time-frequency domains. Lee et al. [169] developed an automatic sleep stage classification system with a mean percentage agreement of 75.52% for diagnosing OSA, using single-channel frontal EEG to classify wake, light sleep, deep sleep, and REM sleep. In order to achieve sleep stage classification, some machine learning-based methods [170, 171] have been introduced in sleep stage classifications, e.g., Support Vector Machine (SVM) [172] and Random Forest (RF) [173]. However, these methods have some limitations, such as the need to observe each PSG epoch for extracting features with prior knowledge. For the time being, more studies are focusing on deep learning-based methods. Owing to the availability of high-quality datasets of EEG signals, deep learning-based methods are widely used to extract features from EEG signals for sleep stage classification. In our opinion, the latest deep learning-based methods for sleep stage classification can be split into two categories: non-GCN-based methods and GCN-based methods.

1. **Non-GCN-based Methods**

   More studies are solving the task of sleep stage classification based on RNNs and CNNs. RNNs are commonly used to model the temporal dynamics of EEG signals [174]. In the SeqSleepNet [175], a hierarchical RNN is used to model the sleep stage and achieve accuracy up to 87.1%. In RNN, there are two kinds of the most representative structures, LSTM [137] and Gated Recurrent Unit (GRU) [176]. For example, IITnet [177] is proposed to automatically score sleep stages via bi-directional long short-term memory (BiLSTM). However, the problem of gradient disappearance or explosion occurs during RNN training, which makes it difficult to train a deep RNN model. Compared to RNNs, CNNs have high performance in parallel computing. To extract local and global features,

Tsinalis et al. [178] proposed an automatic classification approach for sleep stage scoring based on single-channel EEG. Phan et al. [179] used a simple yet efficient CNN to extract sleep features from EEG signals. In addition, SleepEEGNet [180] employs deep CNNs as the backbone network for sleep stage classification, achieving an accuracy of 84.26 %. Chanbon et al. [181] introduce an end-to-end deep learning approach for sleep stage classification using multivariate and multimodal EEG signals. Furthermore, there are some works that combine CNN with RNN to simultaneously extract spatial and temporal features for sleep stage classification, e.g., DeepSleepNet [182] and TinySleepNet [183]. However, EEGs are non-Euclidean data, which naturally results in CNNs and RNNs being limited in feature extractions. Furthermore, their development potential is further hindered by the enormous parameter overhead.

2. **GCN-based Methods**

The GCN [2] is an advanced neural network structure for processing graph-structured data. Since EEG channels are structured data with temporal relations, each channel can be considered as a node in the graph. For this reason, GCN-based methods have been proven to be more powerful in processing EEGs. The joint analysis of EEG and eye-tracking recordings is raised by Zhang et al. [184], whose strategy is to introduce GCN to fuse features. However, EEG channel signals include the temporal dynamic information of brain activity and the functional dependence between brain regions. To remedy the deficiency of the traditional spatiotemporal prediction model, the ST-GCN [98] is proposed to model spatiotemporal relations and to learn the dynamic EEG for the task of sleep stage classification. For example, GraphSleepNet [185] is proposed to utilize brain spatial features and transition information among sleep stages to achieve more specific performance. However, the dependence on non-adjacent electrodes placed in different brain regions is often overlooked. Since then, Jia et al. [186] propose a multi-view spatial-temporal graph convolutional network (MSTGCN) to extract the most relevant spatial and temporal information with superior perfor-

mance. They introduce spatiotemporal attention to extract temporal and spatial information, respectively. However, this method makes it ineffective to capture the spatiotemporal dependencies on separated attention.

4.1.1 Issues

After summarizing the previous works, there are three shortcomings that need to be solved:

1. Topological connections of electrodes in context are not well captured;

2. These previous methods force GCNs to aggregate features in different channels with the same topology, which limits the upper bound of model performance;

3. Attention weights are not sufficient to summarize long-range spatiotemporal characteristics.

4.1.2 Purpose

In order to address the aforementioned challenges, we propose a combination of dynamic and static ST-GCN with inter-temporal attention blocks for automatic sleep stage classification based on EEG.

4.1.3 Outline

The rest of this chapter is organized as follows: In Section 4.2, we present a series of preparatory works for our study. In Section 4.3, we briefly describe the proposed network framework, including the dynamic and static ST-GCN and the inter-temporal attention block. The dataset used, the experiments, and the experimental results are presented in Section 4.4. Finally, we conclude this work in Section 4.5.

4.2 Preliminaries

A sleep stage network is described as an undirected graph $G = (V, E)$, where $V = \{V_1, V_2, \cdots, V_n\}$ is the collection of N nodes representing electrodes in the brain, and the edge set E represents the connection between nodes captured by an adjacency matrix $A \in N \times N$. A is a matrix composed of 0 and 1 , where 1 represents that the corresponding electrodes are connected, and 0 otherwise. Graph G is made up of a 30-second EEG signal sequence S_t. The sleep feature matrix is the input of G. We define the raw signal sequence as $S = \{S_1, S_2, \cdots, S_m\} \in \mathbb{R}^{m \times Q \times T}$, where m denotes the number of samples, Q means the number of electrodes, and T is the time series length of each sample $S_i \in S\,(i \in \{1, 2, \cdots, m\})$. Inspired by Hyvr$\ddot{a}$inen's work [187], we can extract the features of differential entropy (DE) on different frequency bands and define them on each sample feature matrix. Therefore, we can obtain a feature matrix at each sample i, denoting the F_{de} features of the nodes N.

$$X_i = \left(x_1^i, x_2^i, \cdots, x_N^i\right)^T \in \mathbb{R}^{N \times F_{de}} \tag{4.1}$$

Therein, $x_n^i \in \mathbb{R}^{F_{de}}\,(n \in \{1, 2, \cdots, N\})$ denotes the F_{de} features of electrode node n at sample *i*. The objective of our study is to establish a mapping relationship between sleep signals and sleep stages using a spatiotemporal neural graph network. The issue of the sleep stage is described as follows:

$$\mathbb{C} = (X_1, X_{1+d}, \cdots, X_{1+kd}) \in \mathbb{R}^{N \times F_{de} \times T_n} \tag{4.2}$$

The given Equation (4.2) can identify the current sleep stage S. Therein, $\mathbb{C}$ denotes the temporal context of X_{1+kd}, S denotes the sleep stage class label defined by X_{1+kd}, T_n indicates the length of sleep stage networks, d denotes the temporal context coefficient, and k is the number of intercepted time segments in a continuous EEG signal.

4.3 Methods

In this section, we introduce the components of our proposed network of sleep stage classification in detail.

4.3.1 Network Architecture

Figure 4.1 illustrates our network architecture. Inspired by ST-GCN [98], we construct the network of sleep stage classification by nine serial connected ST-GCN modules that can extract more detailed feature information. The ST-GCN module contains a sequential execution of a GCN block and a TCN block. The TCN block is a one-dimensional CNN used for sequence modeling tasks. The GCN block and the TCN block in GCN aggregate features along the spatial dimension and the temporal dimension, respectively. Each ST-GCN module is followed by an attention block (ATT). The function of the ATT block allows the network architecture to pay more attention to important features of the sleep stage, thus better capturing spatiotemporal relations. As far as we know, this is the first attempt to introduce attention enhancement and spatiotemporal separated feature extraction together for sleep stage classification using EEGs. Each module is presented separately in the following subsections.

4.3.2 Graph Convolutional Network Module

In our work, we construct a spatiotemporal graph with the electrodes in the brain as graph nodes and natural connections in different brain region electrodes and time as graph edges. In sleep stage classification tasks, it is important that we model the spatial dependencies in the sleep staging network. GCN is able to effectively extract key point information from the spatiotemporal graph. To capture the dependency created by the topological structures of the electrodes in the context, the layer-wise update rule of GCNs may be implemented to features at time T on sleep inputs defined by features X and the graph structure $\tilde{A}$, as follows:

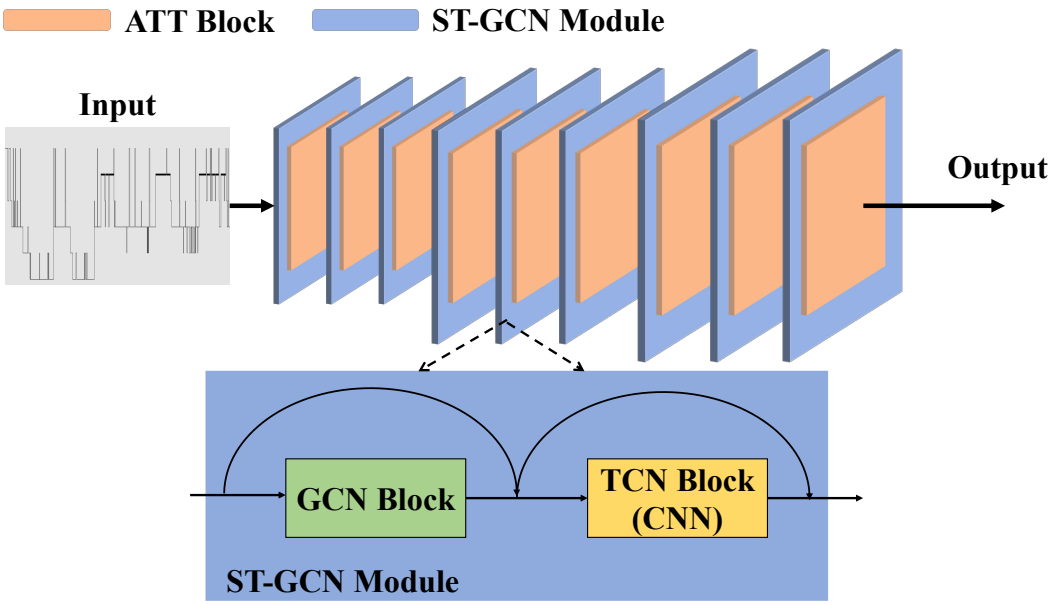


Figure 4.1: **The proposed network architecture for sleep stage classification. The network consists of nine ST-GCN modules, each followed by an attention (ATT) block. Each ST-GCN module contains a GCN block followed by a TCN block. The numbers of output channels for ST-GCN modules are 66, 66, 66, 132, 132, 132, 264, 264, 264.**

$$X_T^{l+1} = \lambda\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}X_T^{(l)}\mu^{(l)}\right) \tag{4.3}$$

Therein, $\tilde{D}$ is the diagonal degree matrix of $\tilde{A}$, and the sleep graph with self-loops $\tilde{A} = A + I$ consists of an adjacency matrix A and an identity matrix I. This allows $\tilde{A}$ to preserve the identity features. The $\lambda(\cdot)$ is an activation function and the μ denotes the weight matrix. Moreover, $\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ can be conceived as an approximate spatial mean feature aggregation from the immediate neighborhood followed by an activated linear layer.

In static methods, $\tilde{A}$ is defined manually or set as a trainable parameter. The topology is predefined according to the structure and is fixed in both the training and testing phases. Notably, these methods have some limitations, such as the need for prior knowledge and the inability to construct dynamic graph topologies. To overcome these limitations, the model is usually required to be generated adaptively depending on the input sample. Therefore, a dynamic ST-GCN [185] is proposed that defines a non-negative function to represent the connection relationship between electrode nodes N_i and N_j

based on the input feature matrix. From this effect, the dynamic adjacency matrix is more powerful since it can be dynamically adapted during the training process and has a stronger generalization ability compared to static methods due to the dynamic topologies. Although the use of dynamic topologies leads to good performance, the original structural information is often discarded. Therefore, we propose a combination of dynamic and static GCN that incorporates contextual features of all brain regions to learn correlations between arbitrary pairs of points.

In the static branch, we use the physical graph G_p from the physical connections of the electrode structure and the parameterized mask G_m is used to pay attention to the physical graph G_p. The static topology information of the electrode structure is extracted in the static branch, which has been shown to be useful for the final prediction. The output of the static branch can be shown as follows:

$$Output_{static} = \lambda \left(G_p + G_m \right) X_T^{(l)} \mu^{(l)} \tag{4.4}$$

In the dynamic branch, the predicted dynamic graph G_d is used as input. The output of the dynamic branch extracts the global context-enriched topology of the electrode structure. We represent the output of the dynamic branch as:

$$Output_{dynamic} = G_d X_T^{(l)} \mu^{(l)'} \tag{4.5}$$

Therein, the learnable kernel $\mu^{(l)'}$ is not shared between the static and dynamic branches. Moreover, we fuse static and context-enriched topology features extracted by the static and dynamic branches using a weighted summation method. It can be expressed as:

$$Output = (1 - \phi)\, Output_{dynamic} + \phi Output_{static} \tag{4.6}$$

where ϕ goes from 0 to 1, which is a balance between the output of the static and dynamic branches.

4.3.3 Multi-scale CNN Module

Temporal modeling is essential to sleep stage classification as well. Many studies [188–190] show that RNNs achieve great performance in temporal modeling tasks. However, the main shortcomings of RNNs are time, cost, and its inability to retain long-term memory. Namely, RNNs cannot perform massively parallel processings like CNNs. TCN [191], as a temporal variant of CNN, has promising performance in time series forecasting. Since sleep stage classification is time-dependent, TCN is used to capture dependencies between sleep stages for achieving sleep stage classification. Multi-scale convolutional neural networks can adaptively fuse multi-scale temporal features extracted by different scale convolution kernels. Thus, they can better model temporal topological features.

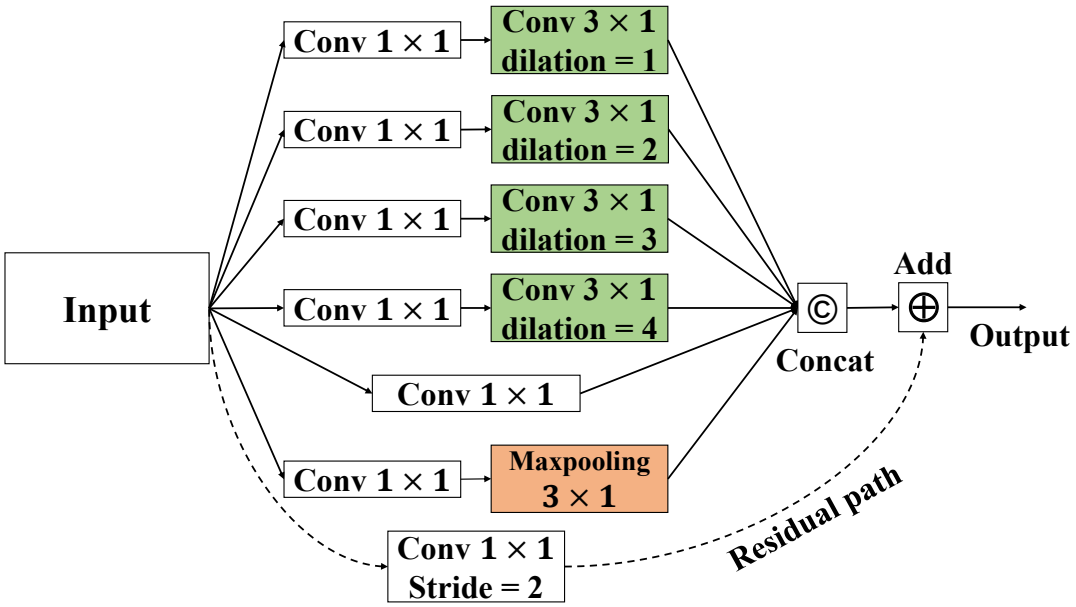


Figure 4.2: **Multi-scale convolutional neural network architecture.**

In order to achieve temporal modeling, many previous studies [85, 106, 108] have used temporal convolutions with a fixed kernel size $k_t \times 1$ throughout the architecture. As a natural extension to the multi-scale spatial aggregation, we used multi-scale learning to improve vanilla temporal convolutional layers, as shown in Figure 4.2. To reduce the computational costs incurred by the extra branches, we introduce the idea of a bottleneck design [192], set the kernel size to 3×1, and employ different dilation

factors [193] instead of larger kernels for larger receptive fields to construct a multi-scale time-series layer. Specifically, seven temporal convolution branches are arranged in parallel. Each branch uses a bottleneck structure (i.e., 1×1 convolution) to reduce the number of feature channels and the calculation amount, thus accelerating the training speed and model inference. Moreover, as the input passes forward, the functions of distinct branches diverge, which can be divided into the following four types.

- Multi-scale temporal feature extraction: In the four temporal convolution branches, each branch consists of 3×1 temporal convolutions. Each 3×1 temporal convolution uses different dilations to obtain multi-scale temporal receptive fields.

- Feature processing within the current frame: This second type only has a temporal convolution with the kernel size 1×1 to concentrate features within a single frame.

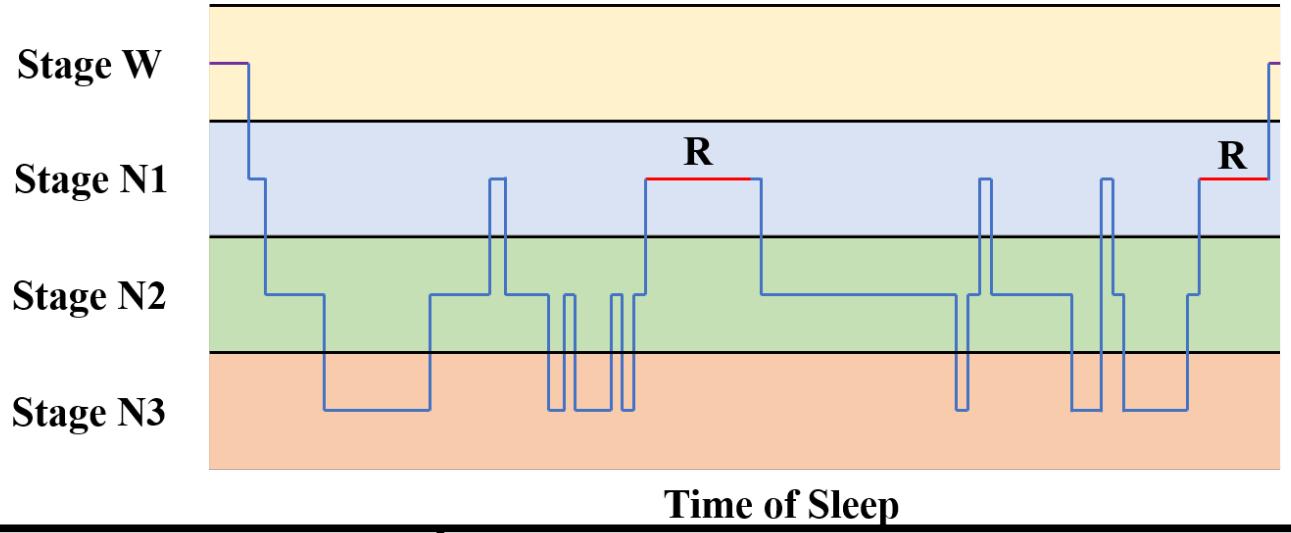
- Emphasizing the most salient information within the consecutive frames: The third type is to be followed by a 3×1 max-pooling layer to draw the most important features.

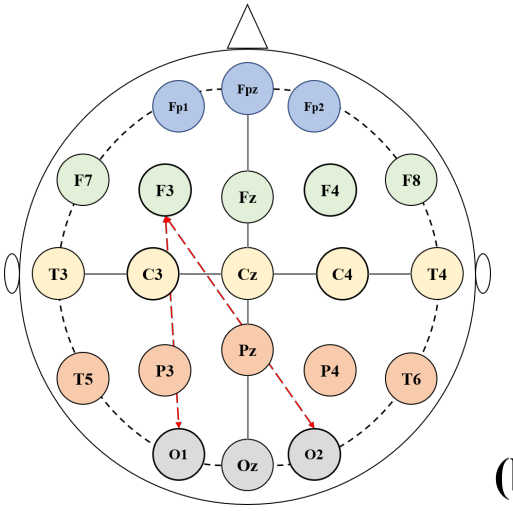
- Gradient preservation: To preserve gradients during back-propagation, we add a residual path in the final type.

Finally, we use residual connections [194] to facilitate training.

4.3.4 Inter-Temporal Attention

Most existing approaches [98, 108, 109] use graph convolution to extract spatial relations at each time step and 1D convolutional layers to model temporal dynamics. However, these methods make it difficult to obtain the direct information flow across spacetime, and complex regional joint spatiotemporal dependencies are not captured. In other words, the factorized modeling cannot capture the long-range features with precise temporal information. In recent years, attention mechanisms have found





Event	Frequency & Duration	Best Electrode Location	Related Stage
Low Amplitude, Mixed Frequency Activity	4 – 7 Hz	Frontal and Central Lobes	N1
Vertex Shape Waves	$<$ 0.5s	Central Lobe	N1
Sawtooth Waves	2 – 6 Hz	Central Lobe	REM
Transient Muscle Activity	$<$ 0.25s	–	REM
K Complex	1.6 – 4 Hz $\geq$ 0.5s	Frontal Lobes	N2
Sleep Spindle	11 – 16 Hz $\geq$ 0.5s	Central Lobes	N2, N3

(c)

Figure 4.3: **(a) An example of a profile of the sleep stages; (b) EEG electrode placement in the 10–20 system, and the F, T, C, P, and O denote frontal, temporal, central, parietal, and occipital lobe placements, respectively; (c) EEG waves and events during sleep [169].**

wide application in various classification tasks, which have made remarkable achievements [73, 195, 196]. The essence of attention mechanisms is to select the relatively critical information from the input. In our work, we consider the spatiotemporal relation of the EEG data and the stability of the learned representations for different sleep stage sequences. For example, in sleep stages R and N_1, the topological features of adjacent electrodes are similar, as shown in Figure 4.3(a) and Figure 4.3(c). To extract strongly distinctive features, there is a need for long-range dependencies in time and precise temporal information in space. In the spatial dimension, the shorter the path length, the more efficient the information transfer between the two electrodes. We pass the relevant features of the distant (informative) electrode to the target electrodes with much higher weights. An example is given in Figure 4.3(b). The feature weights of electrode F_3 are passed to electrodes O_1 or O_2, which can pay attention to important features of distant electrodes in classifying similar sleep stages and better achieve sleep stage classification. Moreover, each electrode is expressed by a time series. In the temporal dimension, there are similarities among neighboring sleep stages, and we attend to important time steps of each electrode. Therefore, inter-temporal attention is introduced to capture the spatial and temporal correlations in the sleep stage classification network.

The classification tasks introduce attention mechanisms to improve the classification effects, which are mainly implemented by a multi-layer perception (MLP), such as the SENet structure [126]. These modules are usually executed independently for each channel or spatial dimension, while other dimensions are globally averaged into a single unit. Since there is a strong link between spatial and temporal information based on GCN in sleep stage classification. It is clear that features separated from frames and electrodes are sub-optimal for weighting the importance of electrodes in different sleep stages, owing to the fact that the spatiotemporal relations are ignored.

We separately consider that the frame and electrode are sub-optimal for weighting the importance of the electrode structure in the sleep stage classification. As an application of coordinate attention [197] for sleep stage classification, we propose an

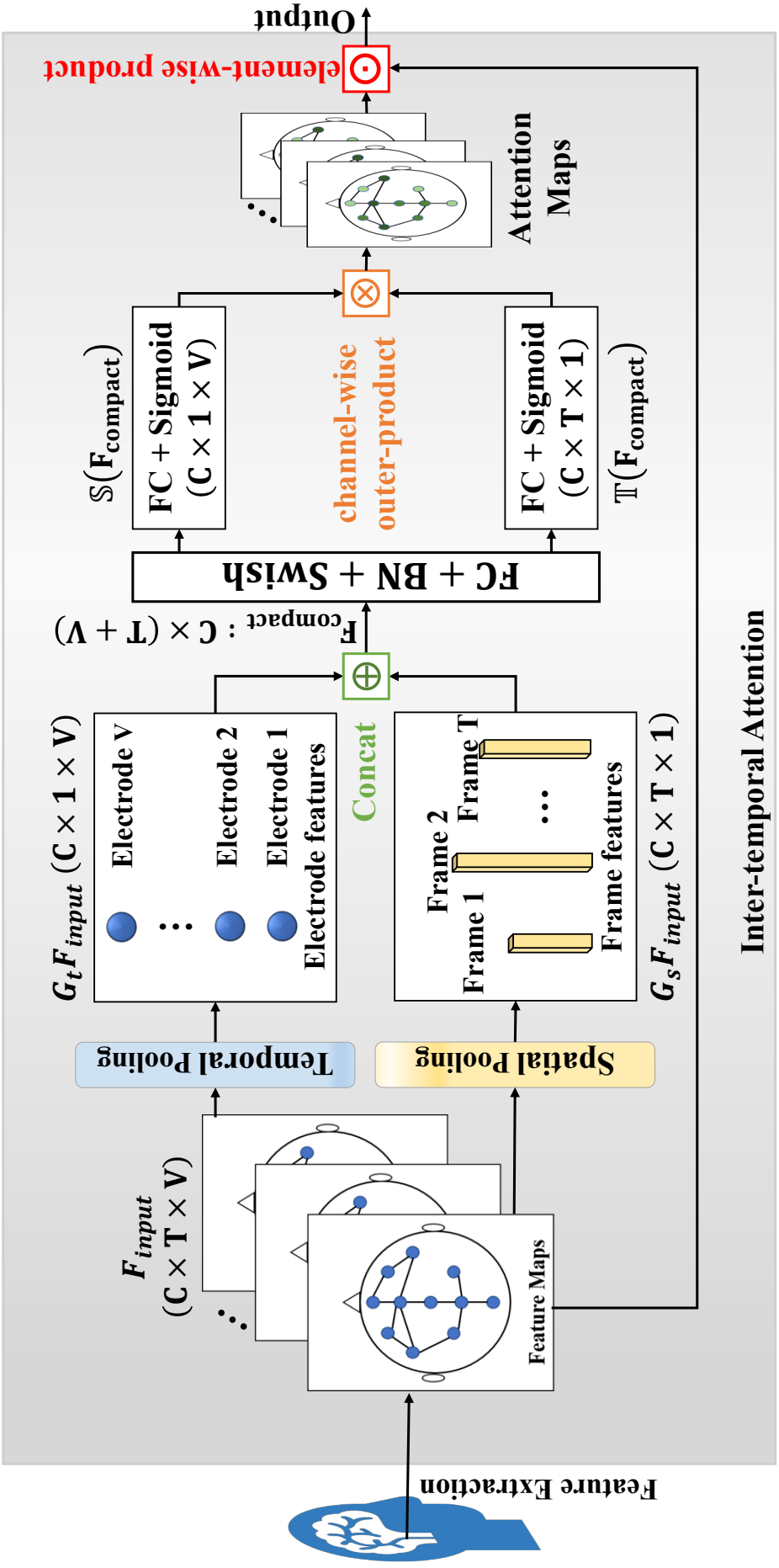


Figure 4.4: **The overview of the inter-temporal attention block. C, T, and V denote the number of input channels, the length of the sequence, and the number of electrodes, respectively. BN denotes the batch normalization.**

inter-temporal attention to enhance the model's ability to extract informative features. It not only identifies the most informative points in certain frames from the whole input sequence, it can also help the network of sleep stage classification to capture richer features. Figure 4.4 is the overview diagram of the inter-temporal attention block. We present the details of an attention block in detail.

- We used a sequence of EEG signals as input, a sequence of EEG signals consists of T number of frames. Each frame consisted of sleep information with dimension $C \times V$, where V is the number of electrodes and C is the number of channels. The input features (F_{input}) were passed through temporal pooling (G_t) and spatial pooling (G_s), respectively. After the operation of pooling, we aggregated the information in the frame- and electrode- dimensions, yielding two sets of feature maps with temporal- and spatial-aware characteristics, the electrode features ($G_t F_{input}$), and the frame features ($G_s F_{input}$).

- We used the concatenation ($\oplus$) operation to obtain the pooled feature vectors ($F_{compact}$), and used the FC layer to obtain the compact information. The activation function Swish (η) [198] is utilized in this FC layer.

- We used two relatively independent FC layers to recover the electrode features and the frame features into the same shape as the input separately. Then, the sigmoid activation function (τ) is applied to the updated tensor. Hence, we can obtain two sets of attention scores, which are from the frame dimensions and the electrode dimensions, respectively. We used the attention scores to reweight the raw feature maps in frame- and electrode- dimensions. Namely, $\mathbb{T}(F_{compact})$ and $\mathbb{S}(F_{compact})$ denote the transfer matrix of the frame and electrode, respectively. In two independent FC layers, we multiplied the obtained attention scores for frame dimensions and electrode dimensions by the channel-wise outer-product ($\otimes$).

- An element-wise product ($\odot$) was performed, resulting in output feature maps (F_{output}). The results of the multiplication could be considered as the attention scores for each electrode in the whole sleep cycle.

The inter-temporal attention module can be explained concisely and intuitively with the following two equations:

$$F_{compact} = \eta\left(MLP \cdot \left(G_t F_{input} \oplus G_s F_{input}\right)\right) \tag{4.7}$$

$$F_{output} = F_{input} \odot \left(\tau\left(\mathbb{T}\left(F_{compact}\right) \otimes \mathbb{S}\left(F_{compact}\right)\right)\right) \tag{4.8}$$

To extract the most noteworthy information from the EEG signal sequence, we perform the max pooling operation under the frame- and electrode- dimensions, respectively. The max pooling plays a similar role as the attention mechanism, the maximum weight of the two dimensions can be selected by this operation. Then, the two groups of the obtained feature maps are concatenated, as shown in Figure 4.5(a). We use the fully connected layer to squeeze the dimensions of the concatenated feature map. Thus, we obtain a continuous feature mapping for our subsequent extraction of the different dimensions of feature attention. After the split operation, two sets of attention scores for the frame dimension and the electrode dimension can be obtained, respectively. What we need is a relationship of attention across time and space, the attention scores of frames and electrodes are multiplied by a channel-wise outer-product, as shown in Figure 4.5(b). Moreover, the result can be seen as the attention scores for each electrode in the whole EEG signal sequence. The attention score is a trainable inter-temporal signal. The joint spatiotemporal attention weight can be seen as the interaction of temporal attention weight and spatial attention weight, and we aggregate the temporal attention branch on the left and the spatial attention branch on the right, as shown in Figure 4.5(c). Finally, we assign the generated spatiotemporal attention weights to the feature maps to obtain the attention responses across space and time. The most informative frames and electrodes can be more accurately located using the attention block, which helps the model to better complete sleep stage classification. As far as we know, this is the first time that inter-temporal attention blocks are introduced for automatic sleep stage classification.

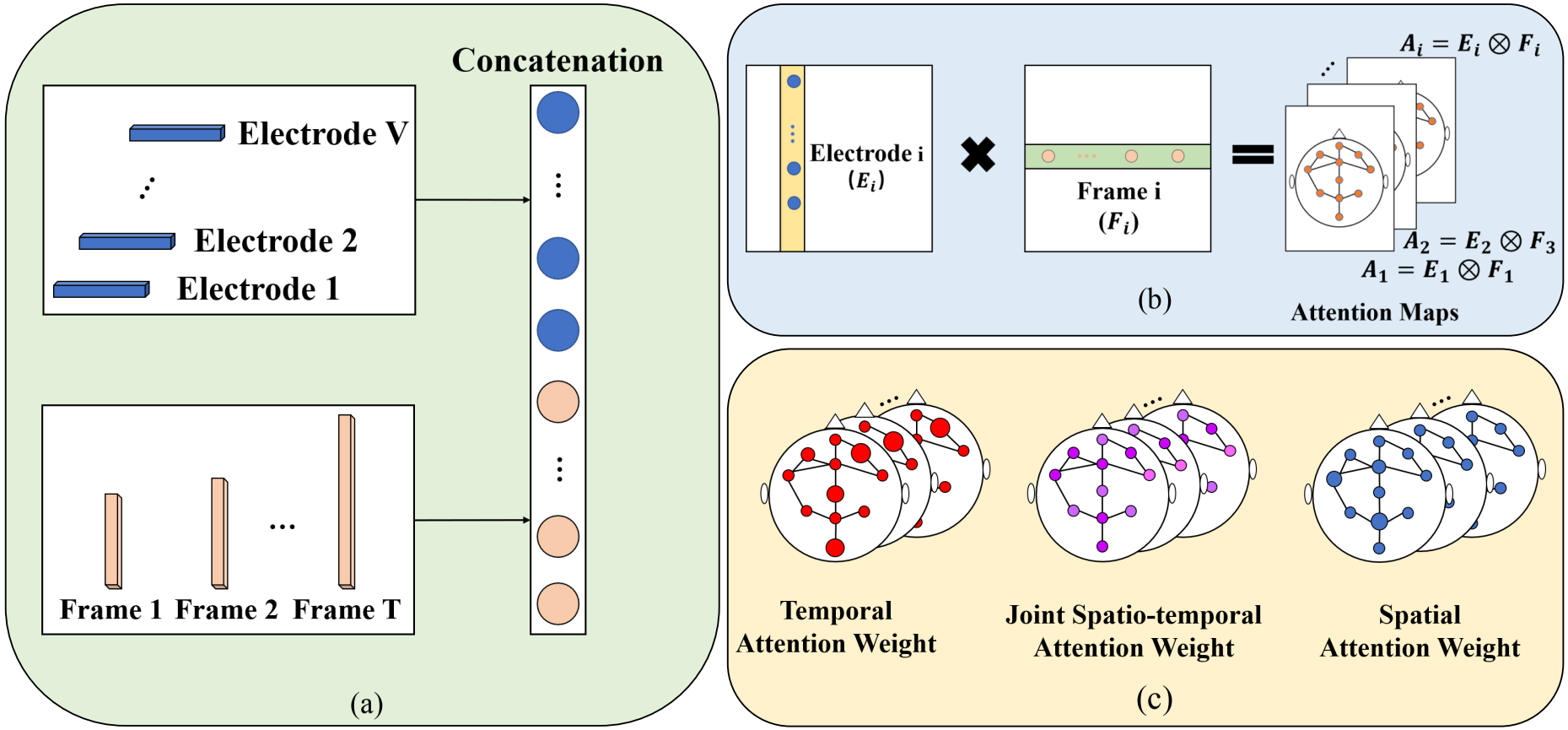


Figure 4.5: **The details of our introduced inter-temporal attention block. (a) The pooled temporal and spatial feature vectors are concatenated; (b) outer product multiplication of frame- and electrode-matrices. Each electrode and the corresponding frame are multiplied with each other to product matrices A, attention maps; (c) example of obtaining the joint spatiotemporal attention weight. The inter-temporal attention blocks capture long-range features with precise temporal information.**

4.4 Results

In this section, we evaluate the performance of the proposed method using two publicly available datasets. A detailed description of the ISRUC-SLEEP dataset, the Sleep-EDF-39 dataset, and the experimental setups can be given in the first two subsections. Then, we report the results of our proposed model compared to the other state-of-the-art models on the same dataset.

4.4.1 Dataset and Experimental Settings

To evaluate the performance of our method, we use the two publicly available datasets in this study: Sleep-EDF-39 dataset [199, 200] and ISRUC-SLEEP dataset [201], which are the most widely used open-source datasets for state-of-the-art methods of sleep stage classification.

4.4.1.1 Sleep-EDF-39 dataset

The Sleep-EDF-39 dataset records the EEG of 20 healthy Caucasian male and female subjects (ages 28.7 $\pm$ 2.9) without medication, and each EEG is sampled at 100 HZ from F_{pz}–C_z and P_z–O_z electrode locations. The EEG recording is manually classified into eight patterns (Wake, S_1, S_2, S_3, S_4, REM, movement, and unknown) according to the scoring rules of R&K [44]. In our experiment, we combine the S_3 and S_4 stages into one stage N_3 according to the AASM manual [45]. As the EEG is recorded over a long period of time, the stages movement and unknown are recorded at the beginning and end of each recording, when the subjects are awake. Therefore, movements (and unknown) are not used for sleep stage classification. Consequently, we obtain a dataset with five classes, including W (Wake), N_1 (S_1), N_2 (S_2), N_3 (S_3 + S_4) and R (REM). We use the 30-min EEG before and after the sleep period as experimental data.

4.4.1.2 ISRUC-SLEEP Dataset

The ISRUC-SLEEP dataset from the Portuguese Foundation for Science and Technology (PFST) has three subgroups, with each subgroup recording the EEGs of 100 participants, 8 participants, and 10 participants, respectively. In order to compare healthy subjects with patients suffering from sleep disorders, we used the subgroup III as the experimental dataset in our study; the EEG recordings of nine healthy male subjects and one healthy female subject aged between 30 and 58 years. Moreover, each EEG recording contained six EEG channels (i.e., C_3–A_2, C_4–A_1, F_3–A_2, F_4–A_1, O_1–A_2, and O_2–A_1) and is sampled at 200 Hz. The EEG recordings were visually scored by a human expert. According to the AASM manual [2], there were five classes in this dataset, including W (Wake), N_1, N_2, N_3, and R (REM). Table 4.1 shows the number of sleep stages in two different datasets.

Table 4.1: Details of the number of sleep stages in the subgroup III of the ISRUC-SLEEP dataset and Sleep-EDF-39 dataset.

Dataset	W	N_1	N_2	N_3	R	Total
Sleep-EDF-39	7927	2804	17,799	5703	7717	41,950
ISRUC-SLEEP	1817	1248	2678	2035	1111	8889

4.4.2 Experimental Settings

In our experiment, we respectively use the 20-fold cross-validation and 10-fold cross-validation to evaluate our method. In each iteration of our methodology, we adopt a leave-one-out approach. The recordings from one subject are designated as the test set, while the recordings from all other subjects are compiled to form the training set. This strategy allows for thorough evaluation and ensures that the model is tested against diverse data sets, contributing to a robust and generalizable system. We implement our model with PyTorch 1.7.1, CUDA 11.4, and Anaconda 4.10.3. The detailed hyperparameters of our experiment are listed in Table 4.2.

Table 4.2: The hyperparameters of our experiment.

Hyperparameters	Value
Optimizer	Adam
Batch size	64
Number of training epochs	120
Learning rate	Initial learning rate is 0.001 and is decayed by 10 at the 30th, 60th, and 90th epoch.
Dropout probability	0.2
Layer number of ST-GCN	9
Reduction ratio	4
Numbers of output channel for ST-GCN	66, 66, 66, 132, 132, 132, 264, 264, 264

4.4.3 The Performance of Sleep Stage Classification

In our study, we use some metrics to evaluate the proposed model [202–204], e.g., the macro-precision, macro-recall, macro-F score, and Cohen's Kappa coefficient. The macro-precision (P_{macro}), macro-recall (R_{macro}), macro-F score ($MF1$), and Cohen's Kappa coefficient (κ) are calculated as follows:

$$ACC = \frac{1}{K}\sum_{i=1}^{K}\left(\frac{TP+TN}{TP+FP+FN+TN}\right)_i \tag{4.9}$$

$$P_{macro} = \frac{1}{K}\sum_{i=1}^{K}\left(\frac{TP}{TP+FP}\right)_i \tag{4.10}$$

$$R_{macro} = \frac{1}{K}\sum_{i=1}^{K}\left(\frac{TP}{TP+FN}\right)_i \tag{4.11}$$

$$MF1 = \frac{1}{K}\sum_{i=1}^{K}\left(\frac{2\times TP}{2\times TP+FN+FP}\right)_i \tag{4.12}$$

$$\kappa = \frac{ACC - p_e}{1 - p_e} \tag{4.13}$$

Therein, TP, FP, and FN, respectively, stand for the true positives, false positives, and false negatives of class i. In our experiment, n represents the number of subjects. In Equation (4.13), ACC is the accuracy of our model, and p_e denotes the hypothetical

probability of chance agreement.

Table 4.3: The confusion matrix of our proposed method on the Sleep-EDF-39 dataset.

		Predicted Stage					
		W	N_1	N_2	N_3	R	**Total**
Actual stage	W	**7371**	214	94	147	101	7927
	N_1	53	**2496**	201	44	10	2804
	N_2	480	552	**16,019**	187	561	17,799
	N_3	147	93	249	**5123**	91	5703
	R	21	103	15	410	**7168**	7717
	Total	8072	3458	16,578	5911	7931	**41,950**

Table 4.4: The confusion matrix of our proposed method on the subgroup III of the ISRUC-SLEEP dataset.

		Predicted Stage					
		W	N_1	N_2	N_3	R	**Total**
Actual stage	W	**1682**	83	37	7	8	1817
	N_1	94	**878**	183	6	87	1248
	N_2	19	179	**2297**	158	25	2678
	N_3	4	3	122	**1905**	1	2035
	R	8	59	37	3	**1004**	1111
	Total	1807	1202	2676	2079	1125	**8889**

Macro-averaged performance obtained with the sleep-EDF-39 dataset and the subgroup III of the ISRUC-SLEEP dataset is shown in Tables 4.3 and 4.4. From Table 4.3, we can calculate that the macro-precision, macro-recall, and macro-F score are 87.4%, 90.9%, and 89.0%, respectively. From the Table 4.4, the macro-precision, macro-recall, and macro-F score are 86.6%, 86.5%, and 86.5%, respectively. In two different datasets, we obtain an accuracy of 91.0 % and 87.4 %, respectively. Cohen's kappa coefficients are 0.88 and 0.84, which is considered ideal as it outperforms the standard of 0.8 [203]. To validate the effect of introducing the ATT blocks, we use a 20-fold cross-validation on the Sleep-EDF-39 dataset and a 10-fold cross-validation on the subgroup III of the ISRUC-SLEEP dataset. The results of the comparisons are described in Figure 4.6. Figure 4.6 presents that the model with the ATT blocks performed better than the model without the ATT blocks in terms of overall accuracy and F1-score for each sleep stage. The performance has been significantly improved.

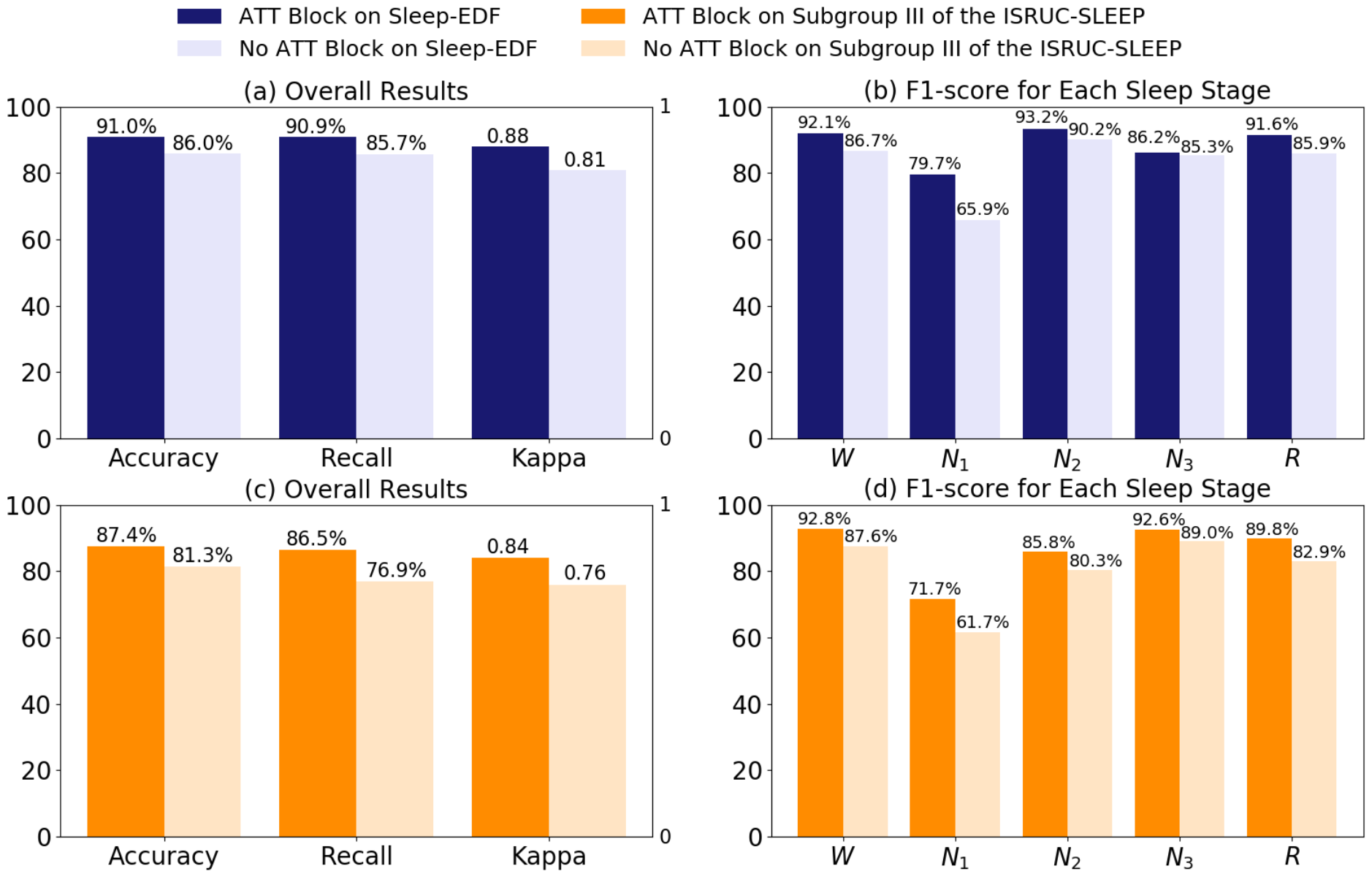


Figure 4.6: **The comparison result of introducing ATT blocks and no ATT blocks. We employ the Sleep-EDF-39 dataset to obtain the comparison results, as shown in sub-figure (a) and sub-figure (b). The sub-figure (c) and sub-figure (d) present the performance comparison of introducing ATT blocks and no ATT blocks on the subgroup III of the ISRUC-SLEEP dataset. Obviously, the model with ATT blocks yields the best results in terms of all kinds of measuring metrics.**

4.4.4 Comparisons with State-of-the-Art Models

To verify the superiority of our proposed model, we compare it with state-of-the-art models on the Sleep-EDF-39 dataset and the subgroup III of the ISRUC-SLEEP dataset. We use the same experimental settings to train all models. Compared to other baseline methods, our model outperforms significantly better than the state-of-the-art methods, as can be seen in Tables 4.5 and 4.6. First, we consider previous works that utilize RNN and CNN to extract spatial or temporal features for sleep stage classification. These non-GCN-based methods use grid data as input to high accuracy. However, EEGs, as non-Euclidean data, can be well processed by powerful GCNs. Therefore, we use two datasets to evaluate the performance of existing GCN-based methods and perform a comparative analysis.

As shown in Tables 4.5 and 4.6, our proposed method presents the best overall performance compared to the state-of-the-art methods. The proposed method achieves the best accuracy (91.0% and 87.4%), the macro-F score (89.0% and 86.5%), and Kappa (0.88 and 0,84) on two datasets. For the subgroup III of the ISRUC-SLEEP dataset, the proposed method provides the highest accuracy for each sleep stage. For the Sleep-EDF-39 dataset, our method achieves the highest accuracy for each sleep stage except for N_3 stage (sub-optimal). For N_1 stage, Tables 4.5 and 4.6 show that the classification effect for N_1 stage on the two datasets is not as ideal as for the other sleep stages. It can be explained by two reasons. First, a number of samples in N_1 stage belong to the sleep transition period [205], thus the N_1 stage is misclassified into other stages. Second, the N_1 stage occupies a small proportion of the dataset. In particular, in the Sleep-EDF-39 dataset, the proportion of N_1 stage is only 6.7%.

4.5 Discussion

Sleep disorders are highly prevalent in the world. Especially in the United States, nearly 25% of adults suffer from sleep disorders [208]. Sleep disorders not only affect the quality of life, but also lead to health problems, such as heart disease and stroke.

Table 4.5: Comparison between our proposed method and the other state-of-the-art methods on the Sleep-EDF-39 dataset across overall performance and F1-score for each sleep stage. The numbers in bold indicate the highest performance metrics of all methods and the underlined result is the sub-optimal result.

	Performance of Quality Assessment			**Global F1-Score for Sleep Stage (%)**				
Method (Year)	**Accuracy (%)**	**Macro-F Score (%)**	**Kappa**	**W**	N_1	N_2	N_3	**R**
Non-GCN-Based Methods								
Tsinalis et al. [178] (2016)	74.8	69.8	–	65.4	43.7	80.6	84.9	74.5
Tsinalis et al. [168] (2016)	78.9	73.7	–	71.6	47.0	84.6	84.0	81.4
DeepSleepNet [182] (2017)	82.0	76.9	0.76	84.7	46.6	85.9	84.8	82.4
SeqSleepNet [175] (2017)	81.2	74.6	0.73	74.1	46.9	86.9	81.2	83.8
Phan et al. [179] (2018)	82.3	74.7	0.75	77.3	40.5	87.4	86.0	82.3
IITNet [177] (2019)	84.0	77.7	0.78	87.9	44.7	88.0	85.7	82.1
SleepEEGNet [180] (2019)	84.3	79.7	0.79	89.2	52.2	86.8	85.1	85.0
TinySleepNet [183] (2020)	85.4	80.5	0.80	<u>90.1</u>	51.4	88.5	**88.3**	84.3
GCN-Based Methods								
GraphSleepNet [185] (2021)	84.2	81.0	0.79	83.2	69.0	88.4	74.9	89.6
Jia et al. [186] (2021)	<u>86.4</u>	<u>84.1</u>	<u>0.82</u>	85.5	<u>75.3</u>	<u>89.8</u>	80.4	<u>89.3</u>
Our proposed method	**91.0**	**89.0**	**0.88**	**92.1**	**79.7**	**93.2**	<u>88.2</u>	**91.6**

Table 4.6: Comparison between our proposed method and the other state-of-the-art methods on subgroup III of ISRUC-SLEEP dataset across overall performance and F1-score for each sleep stage. The numbers in bold indicate the highest performance metrics of all methods and the underlined result is the sub-optimal result.

	Performance of Quality Assessment			**Global F1-Score for Sleep Stages (%)**				
Study (Year)	**Accuracy (%)**	**Macro-F Score (%)**	**Kappa**	**W**	**N_1**	**N_2**	**N_3**	**R**
Non-GCN-Based Methods								
Memar et al. [173] (2017)	72.9	70.8	0.65	85.8	47.3	70.4	80.9	69.9
Dong et al. [170] (2017)	77.9	75.8	0.71	86.0	46.9	76.0	87.5	82.8
DeepSleepNet [182] (2017)	78.8	77.9	0.73	88.7	60.2	74.6	85.8	80.2
RotSVM [172] (2018)	73.3	72.1	0.66	86.8	52.3	69.9	78.6	73.1
Phan et al. [179] (2018)	78.9	76.3	0.73	83.6	43.9	79.3	87.9	<u>86.7</u>
Chambon et al. [181] (2018)	78.1	76.8	0.72	87.0	55.0	76.0	85.1	80.9
Ghimatgar et al. [206] (2019)	75.7	73.5	0.69	85.0	49.4	75.4	83.1	74.8
Shen et al. [207](2020)	81.7	80.2	0.76	89.1	<u>62.5</u>	80.4	86.5	82.4
GCN-Based Methods								
GraphSleepNet [185](2021)	79.9	78.7	0.74	87.8	57.4	77.6	86.4	84.1
Jia et al. [186](2021)	<u>82.1</u>	<u>80.8</u>	<u>0.77</u>	<u>89.4</u>	59.6	<u>80.6</u>	<u>89.0</u>	85.6
Our proposed method	**87.4**	**86.5**	**0.84**	**92.8**	**71.7**	**85.8**	**92.6**	**89.8**

For people with sleep disorders to obtain adequate sleep, they may require the help of an appropriate method for sleep stage classification. In this work, we use a combination of dynamic and static ST-GCN with inter-temporal attention blocks to automatically classify sleep stages. We first consider that the distribution of brain electrodes is characteristic of non-Euclidean data. After the addition of ATT blocks, the sleep stage classification network achieves better performance. This confirms that spatial and temporal correlations play an important role in the sleep stage classification. The obtained results suggest that our method is promising in detecting new abnormalities in sleep and continuously improving our understanding of sleep mechanisms.The NREM stages are divided into three sleep stages (N_1, N_2, and N_3) and are associated with the depth of sleep. Research shows that the stage N_3 may affect the ability to learn new information and memory retention [209]. In simple terms, N_3 is the deepest sleep stage, which has the strongest repair function. Tafaro et al. [210] report a positive relationship between sleep quality and survival in centenarians. From our experiment, the proposed method shows excellent performance in classifying the stage N_3 compared with stages N_1 and N_2. Therefore, accurate detection of the stage N_3 provides an aid to long-term care, health and welfare services for the elderly. A study [211] shows that patients with REM_{OSA} in REM sleep had a significantly more collapsed airway and better ventilatory control stability compared with NREM sleep. Moreover, as it is suggested that the increased proportion of N_3 stage may reveal a lower severity of OSA [212], our method can be used as an ancillary treatment.

There are several challenges in the broader context of sleep stage classification. Firstly, accurately detecting stage N_1 is challenging since it is a transitional phase between wakefulness and sleep. The system should be improved for the diagnosis of sleep fragmentation, such as obstructive sleep apnea. Secondly, the dataset quality is often compromised due to human errors. Given that sleep scoring is typically done by experts, it's common for similar sleep stages to be mislabeled. Consequently, a significant issue for many sleep stage classification networks is utilizing high-quality datasets for training. In response to these challenges, future developments aim to create a sleep

stage system that emulates more human-like performance, thereby enhancing accuracy and reliability in sleep stage classification.

Chapter 5

4s-SleepGCN: Four-Stream Graph Convolutional Networks for Sleep Stage Classification

In this chapter, we present a multimodal physiological signals-based approach for sleep stage classification. Sleep stage classification serves as a critical basis for assessing sleep quality and diagnosing sleep disorders in clinical practice. Most existing methods rely solely on a single channel for sleep stage classification, thereby neglecting the complementary nature of multimodal electrophysiological signal characteristics. In contrast, the current multi-stream sleep staging network primarily utilizes EOG and EEG signals as inputs and efficiently fuses the extracted multimodal features. However, the importance of motion information in electrophysiological signals is rarely investigated, which could improve classification performance. Moreover, recent sleep staging models have been plagued by issues of overparameterization and suboptimal classification accuracy. Moreover, EOG and EEG are non-Euclidean graph-structured data that can be effectively handled by graph convolutional networks. To address the aforementioned issues, we propose an efficient graph-based multi-stream model named 4s-SleepGCN, which combines biological signal features to classify sleep stages. In each single-stream model, the positional relationship of the modal sequences is incor-

porated into the proposed model to enhance the feature representation for sleep stage classification. On this basis, graph convolutions are utilized to capture spatial features, while multi-scale temporal convolutions are employed to model temporal dynamics and extract more discriminative contextual temporal features. The EEG signal, EOG signal, and corresponding motion information are separately fed into the single-stream model comprising our 4s-SleepGCN. Experimental results show that the proposed 4s-SleepGCN achieves the highest accuracy compared to state-of-the-art methods in both the Sleep-EDF-39 dataset (92.3%) and the Sleep-EDF-153 dataset (85.5%). Additionally, we conduct numerous experiments on two representative datasets that demonstrate the validity of the motion modalities in sleep stage classification. Also, the proposed single-stream network shows higher accuracy (89.2% and 89.8%) in classification while requiring 33% fewer parameters. Our proposed 4s-SleepGCN model serves as a powerful tool to assist sleep experts in assessing sleep quality and diagnosing sleep-related diseases.

5.1 Introduction

Cognitive computing [213], a multifaceted field that combines computer science, cognitive science, and neuroscience, aims to replicate human cognitive processes and develop intelligent systems. This field has a significant application in the analysis and interpretation of sleep patterns. Quality sleep is essential for productive daily life and overall health [214]. However, sleep disorders [215] can drastically reduce sleep quality, leading to various health complications. In particular, these disorders impair physical performance during the day and cognitive functions, such as attention, learning, and memory in the long term [216]. Sleep monitoring systems that incorporate sleep stage scoring, are of crucial importance in sleep medicine. They provide key insights into individual sleep patterns and are essential for the diagnosis and treatment of sleep disorders. Since PSG is the most important test for the diagnosis of sleep disorders through continuous monitoring to understand the patient's condition, many

researchers [23, 217–219] are utilizing sleep monitoring systems based on PSG signals. These systems provide an objective assessment of sleep quality and are of great importance for the prevention and diagnosis of sleep disorders. Thus, PSG signals can play a crucial role in this area of sleep monitoring systems enabled by cognitive computing technologies.

Traditionally, sleep stages are determined by human experts who analyze biological signals recorded during the nocturnal PSG. The AASM has established detailed guidelines for this process that apply to both manual and automated classification methods. However, manual classification of sleep stages is a labor-intensive, time-consuming, and error-prone process, as recent studies have emphasized [46]. In contrast, automatic sleep stage classification has been shown to be a robust alternative, demonstrating both reliability and high accuracy. This method significantly improves the efficiency and accuracy of sleep disorder diagnosis. Its increasing effectiveness and practicality are attracting great interest in the field of sleep research.

Over the last decade, there has been a significant increase in the development of automatic sleep stage classification methods, making the concept of automated sleep scoring more feasible. Various techniques have been employed to capture the distinct patterns of brain wave activity characteristic of different sleep stages, enhancing the accuracy of sleep stage classification. For instance, some earlier conventional approaches [168] utilized hand-engineered features from the time, frequency, and time-frequency domains for this purpose. Additionally, methods based on machine learning [172, 173] have shown impressive performance in identifying sleep stages. Nonetheless, these methods primarily rely on hand-crafted features, meaning that the effectiveness and efficiency of the classification largely depend on the quality of feature engineering and the researchers' depth of understanding of the data. As time has progressed, deep learning techniques have increasingly become the norm in the field of sleep stage classification. Each sleep stage is marked by distinct brain wave activity patterns, which are reflected in the shape of EEG time waveforms used for sleep staging. Further, indications of different sleep stages can be detected in other types of signals, such as ECG or

EMG, which are often recorded simultaneously to enhance the accuracy of sleep stage classification. These advancements in deep learning have paved the way for methods that classify sleep stages based on single-modality data, as well as those utilizing combined multimodal information. Consequently, the most recent deep learning-based methods for sleep stage classification can be categorized into two primary frameworks: the single-channel EEG-based method and the multi-modal physiological signals-based method. These approaches represent a significant leap forward in the field, offering a more sophisticated and nuanced analysis of sleep stages.

5.1.1 Single-channel EEG-based Methods

Given the growing trend in the application of deep learning, recent studies have been focusing on the task of sleep stage classification on EGG signals to achieve outstanding performance, which can be roughly divided into three main approaches, namely RNNs, CNNs, and GCNs. RNNs [220] are considered to be able to model the long-term contextual dependencies of temporal sequences in EEG signals. More recently, specific RNN-based methods [221, 222] that learn sequential features from EEG signals have achieved success in automatic sleep staging. In addition, the LSTM, a representative structure of RNN, has demonstrated great effectiveness and is utilized in IITnet [177] to learn the transition rules among sleep stages. However, due to the long-term dependence of the data on RNNs, the problem of gradient disappearance or explosion is extremely prone to occur, leading to instability in training the model. In contrast, CNNs have better parallelizability and have the ability to directly extract sleep stage transition features from texture images encoded from sleep stage sequences. The CNN-based method proposed by Tsinalis et al. [178] demonstrates the ability to reliably score sleep stages using a single-channel EEG signal. Sors et al. [223] employ CNNs to extract appropriate features directly from raw EEG. Fang et al. [224] design a novel adaptive-boosting-based dual-stream network framework to extract different modalities features of single-channel EEG signals for sleep staging. In addition, a novel CNN framework based on single-channel EEG signals, called SleepEEGNet [180], has been

proposed for sleep stage evaluation using extracted time-invariant features. However, most CNN-based methods struggle to capture temporal dependencies from EEG signals. To address this issue, several integrated systems (i.e., DeepSleepNet [182] and TinySleepNet [183]) have been proposed, which combine CNN and RNN to simultaneously extract features in the spatial and temporal domains, resulting in accurate models for sleep stage discrimination. Considering that EEG electrodes are distributed in a non-Euclidean space, CNNs and RNNs are limited in that the grid data are used as model input and the connection between spatial correlations between electrodes is ignored. GCNs [2] have been shown to be powerful in modeling the topological relationship of EEG electrodes. In this regard, the ST-GCN [98], as one of the most advanced extensions of GCN-based models, has exhibited outstanding performance in sleep stage classification. A quintessential example should be cited that the GraphSleepNet [185] utilizes spatial graph convolutions in conjunction with interleaving temporal convolutions to effectively capture the transition rules among different sleep stages. Furthermore, Jia et al. [186] have developed a novel deep graph neural network named MSTGCN to extract time-varying spatial and temporal features from multi-channel brain signals, using the spatial topological information between brain regions to distinguish different sleep stages. However, these methods overlook the significance of spatiotemporal relations in sleep staging. To address this limitation, Li et al. [119] propose a combination of dynamic and static STGCN, incorporating inter-temporal attention blocks. This approach effectively captures long-range dependencies among different EEG signals and achieves superior performance in sleep stage classification. Despite achieving better performance, single-channel EEG-based methods are frequently limited by the fact that a single fixed physiological signal is suboptimal for distinguishing specific sleep stages.

5.1.2 Multi-modal Physiological Signals-based Methods

The multi-modal fusion strategy aims to integrate diverse media types, capturing complementary information and thereby enhancing the performance and robustness of learning [225, 226]. Sleep staging is a complex dynamic process, where different

Table 5.1: Representative EEG and EOG Characteristics during Different Sleep Stages.

Sleep stages	EEG-characteristics	EOG-characteristics
REM	Low-amplitude, mixed-frequency EEG activity without K complexes or sleep spindles. (Resembles eyes open wake epoch)	Rapid eye movements.
N_1	Low-amplitude, predominantly 4–7 Hz, mixed EGG activity.	Slow, rolling eye movements.
N_2	Sleep spindles: a train of distinct 11–16 Hz waves (most frequently 12–14 Hz) with a duration between 0.5 and 2 seconds. K complex: negative, well-delineated, sharp waveforms immediately followed by a high-voltage slow wave, with a total duration of at least 0.5 seconds.	Either slow eye movements or absence of slow eye movements.
N_3	Delta waves of high amplitude (greater than 75μV) and low frequency (0.5–2 Hz).	None
Wake	Eye-close wakefulness: sinusoidal alpha rhythm (8-13 Hz activity). Eye-open wakefulness: Beta wave(highest frequency and lowest amplitude).	Eye-close wakefulness: slow-rolling eye movements. Eye-open wakefulness: rapid eye movements.

sleep stages are classified based on physiological signals that exhibit varying frequencies and amplitudes at different time periods. Table 5.1 shows representative EEG and EOG characteristics during different sleep stages, based on information from existing studies [227, 228]. In the N_2 and N_3 stages, the EOG waves exhibit a similar pattern, whereas EEG, as an unimodal physiological signal, provides valuable and specific characteristic information, enabling better classification. In contrast, when classifying the REM and N_1 stages, the EEG signal, which lacks some key features, may lead to misclassification. Therefore, the effective identification of different sleep stages requires the integration of different physiological signals. In order to harness the complementary potential of PSG signals, researchers have turned to utilizing multi-modal signals to enhance sleep staging models. For instance, a variation of CNN [229] demonstrates that using multi-channel data achieves better performance compared to single-channel data. Dong et al. [170] apply a combination of DNN and RNN to extract salient features from EEG and EOG signals. Additionally, Andreotti et al. [230] highlight the advantages of incorporating multi-modal PSG signals for sleep stage classification. And the SeqSleepNet [175] achieves an overall classification accuracy of 87.1% based on multi-channel signals by relying solely on a hierarchical RNN. In a similar vein, Chambon et al. [181] use a spatiotemporal CNN model to capture modality-specific information from all multivariate and multi-modal PSG signals. Phan et al. [179] employ a multi-task CNN combining joint classification and prediction framework to identify sleep stages. These methods primarily focus on extracting the features from different PSG signals individually and combining them by concatenation. However, this is not sufficient to model complex relationships between multimodal signals. As a result, recent works have emerged that fully fuse multimodal feature information to showcase the distinct contributions of each modality in identifying specific sleep stages, such as SalientSleepNet [231] and SleepPrintNet [232]. Moreover, Jia et al. [233] design a squeeze-and-excitation network to model the heterogeneity between different modalities. In the latest research, MMASleepNet [234] introduces an effective feature fusion module to capture the relationships among different modalities. MaskSleepNet [235]

effectively combines CNN with an attention mechanism to capture feature information from different PSG signals, leading to a classification accuracy of up to 85.0% on the Sleep-EDF-153 dataset. However, these methods fail to consider coherent features of the PSG signals, such as the speed at which different PSG signals change from frame to frame. Essentially, comprehensive spatial-temporal dependencies may be ineffectively captured.

5.1.3 Issues

After a thorough review of previous studies, we have identified three main limitations that need to be addressed.

1. The majority of existing multichannel-based methods only consider the captured features from the EEG and EOG signals and ignore the signal motion stream, which is not able to obtain more comprehensive features;

2. Current multistream models for sleep staging are typically overparameterized to extract discriminative features from signal sequences, resulting in high model complexity and limiting the development of multichannel-based sleep staging;

3. In current GCN-based approaches to sleep staging, there is a lack of adequate exploration of the semantic information of signal sequences and long-range spatiotemporal dependencies are not well captured.

5.1.4 Purpose

To address the aforementioned limitations, we propose a novel graph-based multistream fusion model called 4s-SleepGCN for automatic sleep staging. Our proposed model simultaneously fuses the features of EEG signals, EOG signals, EEG motion, and EOG motion within a unified GCN framework. Our proposed model provides a better balance between performance and parameter scale than some state-of-the-art models, achieving the highest overall performance on two standard datasets.

5.1.5 Outline

The remainder of this chapter is organized as follows: Section 5.2 elaborates on the proposed 4s-SleepGCN and explains its components in detail. Next, the dataset used and the experimental settings are described in Section 5.3. Meanwhile, Section 5.3 verifies the effectiveness and advantages of the proposed model using two publicly available datasets. In Section 5.4, we discuss our proposed approach formally.

5.2 Methodology

In this section, we propose a multi-stream framework to fuse the spatial information of two different PSG signals (i.e., EEGs and EOGs) and the motion information of their sequences to obtain a powerful sleep staging model. Accordingly, in this section, we present the architecture and components of our proposed network in detail. The proposed network consists of four functional modules: encoder, position embedding, graph convolutional network module, and temporal modeling module. Finally, a multi-stream feature extraction strategy is introduced to promote the sleep stage classification task.

5.2.1 Network Architecture

Inspired by the success of the two-stream framework and graph convolution [109], we design a graph-based multi-stream network to classify sleep stages from different perspectives. In Figure 5.1(a), the PSG data is preprocessed to obtain EEG sequence, EOG sequence, EEG motion, and EOG motion information. Subsequently, the four data are respectively fed into the SleepGCN network to obtain the softmax scores. As described in [236], the weighted average method has been successfully applied in the field of fusing classification results and can further improve the classification results. Therefore, The prediction of sleep stage classification is calculated by the weighted summation method of the four softmax scores. Figure 5.1(b) illustrates the architecture of the SleepGCN. Among them, the input signal sequence is composed of T frames,

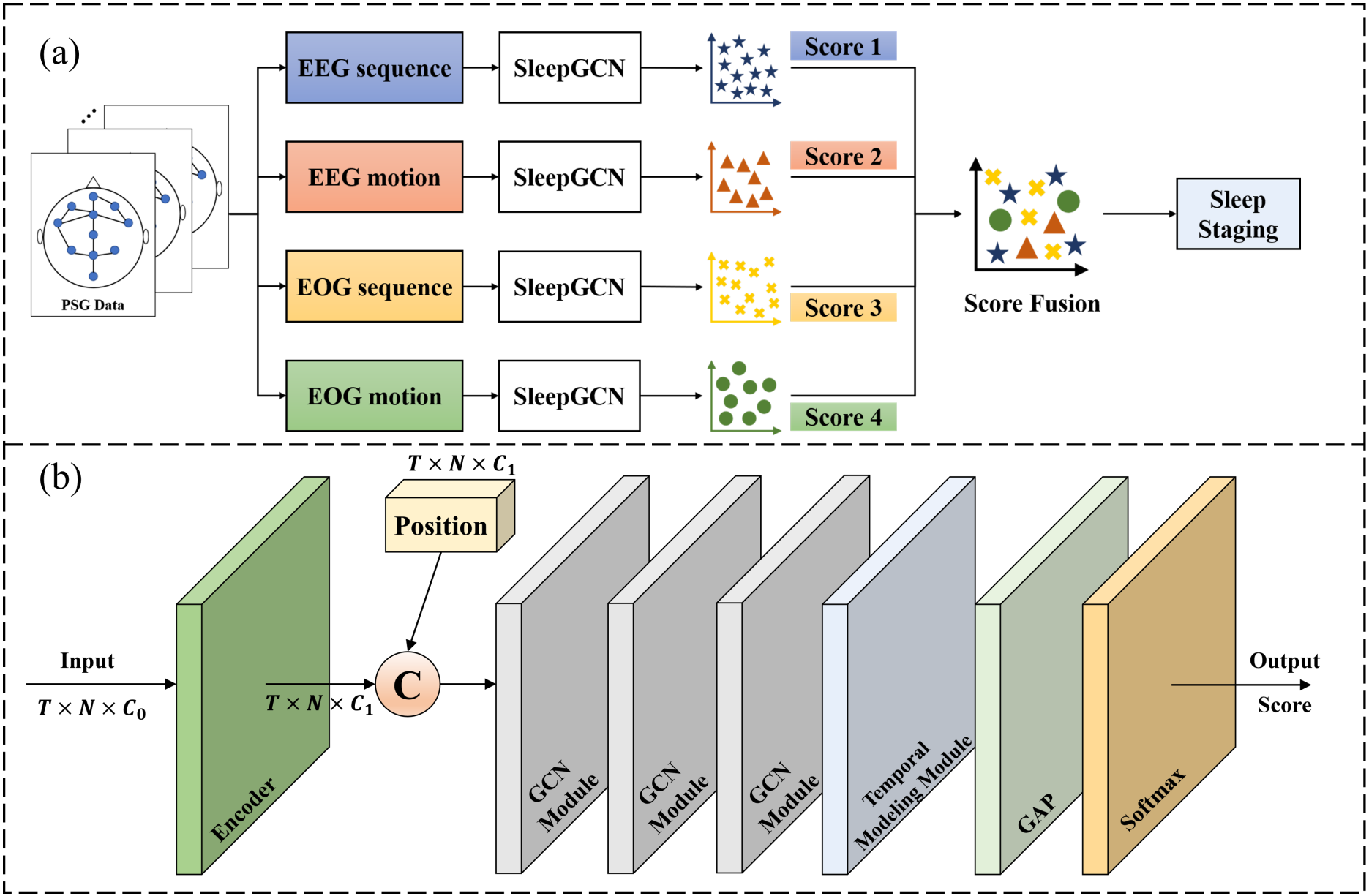


Figure 5.1: **The proposed network architecture for sleep staging. (a) Illustration of the overall architecture of the multi-stream fusion sleep staging network (4s-SleepGCN). (b) Overview of the SleepGCN.**

and the sleep information contained in each frame is composed of the number of electrodes (N) and the number of channels (C_0) with dimensions $C_0 \times N$, which can be represented as an input tensor with the shape of $C_0 \times T \times N$, where C_0 is equal to 3. Then we use two fully connected layers to encode the position to a dimension of 64 (C_1) and then merge it with the position of the same dimension to obtain the new input for 128. The GCN module is adopted to capture long-range spatial dependencies. In order to mitigate the prevalent issue of over-smoothing encountered in most GCN-based models, which has been documented in previous studies [237], we employ ReLU activation functions for each GCN block of our proposed model. By applying activation functions after each GCN block, our classification network can effectively capture complex patterns in the PSG graph data and preserve the expressive power of the node representations, enhancing the model's ability to perform accurate sleep graph classification. The temporal modeling module uses different dilation convolutions to effectively aggregate contextual information. The Global Average Pooling (GAP) layer is introduced to aggregate spatio-temporal features and pool feature maps of distinct samples to a similar size of $1 \times 1 \times 512$. Finally, the softmax layer is used to obtain probabilities for the sleep stage. Each module is presented separately in the following subsections.

5.2.2 Encoder

Since sleep staging based on PSG data can be formulated as a graph modeling problem, the raw PSG sequence of sleep staging can be represented as an undirected graph $G = (V, E)$ with N electrodes and T frames, including a node set $V = \{V_1, V_2, \cdots, V_N\}$ of electrodes N and E is the edge set representing the connection between the electrodes captured by an adjacency matrix $A \in \{0, 1\}^{N \times N}$. A denotes the relationship between the electrodes, where initially $A_{i,j}$ = 1 if there is a functional connection between electrodes i and j, and 0 otherwise. The PSG signal sequence can provide the coordinates of each electrode in graph convolutional networks, which can be described as $X \in \mathbb{R}^{T \times N \times C}$. Therein, N denotes the total number of electrodes in a frame, T is

the number of frames in the raw signal sequence, and C represents the coordinates of all electrodes in the entire frame sequence. We denote all electrode features as a feature set X, which can be represented as a matrix:

$$X = \left\{ X_{n,t} \in \mathbb{R}^{C_0} | n \in N; t \in T \right\} \tag{5.1}$$

where the electrode of type $n = \{1, 2, \cdots N\}$ at time $t = \{1, 2, \cdots T\}$ generates the dimensional feature vector $X_{n,t}$. Our goal is to employ the encoder including two FC layers to encode the original position information into a high-dimensional space, which can be described as follows:

$$X^{'} = ReLU\left(g\left(ReLU\left(\tilde{g} \cdot X + k_1\right) + k_2\right)\right) \in \mathbb{R}^{C_1} \tag{5.2}$$

where $g \in \mathbb{R}^{C \times C_1}$ and $\tilde{g} \in \mathbb{R}^{C_1 \times C_1}$ denotes weight matrices. k_1 and k_2 are the bias vectors. We use the ReLU function as the activation function. In this work, the higher order information by encoding instead of the original position is used as input to improve the ability of personalized expression.

5.2.3 Position Embedding

Position embedding is a widely employed technique for capturing location information within sequences. It has shown successful applications across various domains, with particular effectiveness in natural language processing. Since EEGs and EOGs are time-series data, the sequential relationship between frames affects the meaning of the entire signal. Considering only the coordinate information of the electrodes and the graph structure of the biosignals, it is difficult for the model to capture the sequential relationships between different time steps in the signal, which may result in suboptimal classification performance. Therefore, the absence of the position relationships of sequences could weaken the classification ability of sleep stage models. To address this issue, position embedding is applied in our model to incorporate positional information in the model input, which can better capture the sequential relationships

between different time steps in the sleep signals, leading to improved sleep stage classification performance. Inspired by the previous works [238, 239], two one-hot vectors are applied to characterize the position relations of electrodes and frames. In frame sequences $T = \{T_1, T_2, \cdots, T_w\}$, the w^{th} frame T_w is denoted by a one-point vector, where the w^{th} dimension is set to one and the others are zero. As for the same operation of the frame sequences, we proceed to obtain a one-hot vector as T_w for the electrode sequences. Similar to the encoding of the inputs according to Equation 5.1, the embedding representation in the electrode and frame sequences can be expressed as $N_w^{'} \in \mathbb{R}^{C_1}$ and $T_w^{'} \in \mathbb{R}^{C_1}$, respectively. Subsequently, the embedding vectors in the frame- and electrode- dimensions are fused and concatenated with the original features X'. Finally, the output feature maps $X'' \in \mathbb{R}^{2C_1 \times N \times T}$ can be obtained by the concatenation operation $\frown$, as given in Equation 5.2. Notably, we use the original position as the residual embedding to make the position encoding information explicitly.

$$X'' = \left(N_w^{'} + T_w^{'}\right) \frown X' \tag{5.3}$$

5.2.4 Graph Convolutional Network Module

Indeed, capturing long-range dependencies of PSG sequence data is crucial for sleep stage classification. Inspired by the idea of semantics-guided neural network [238] and non-local block [135], we adopt the GCN module (see Fig. 5.2) to extract correlations between electrodes, thereby capturing rich features of sleep stages from PSG data to achieve sleep staging. More specifically, the similarity between the electrodes in the feature space is used to construct the sleep graph. The long-range weight can be modeled by the pairwise similarity between every two electrodes a^{th} and b^{th} in the same frame T, which is defined as follows:

$$f\left(a, b\right) = \varphi\left(X_a^{''}\right)^T \lambda\left(X_b^{''}\right) \tag{5.4}$$

where φ and λ represent two transformations of the original features. Since the

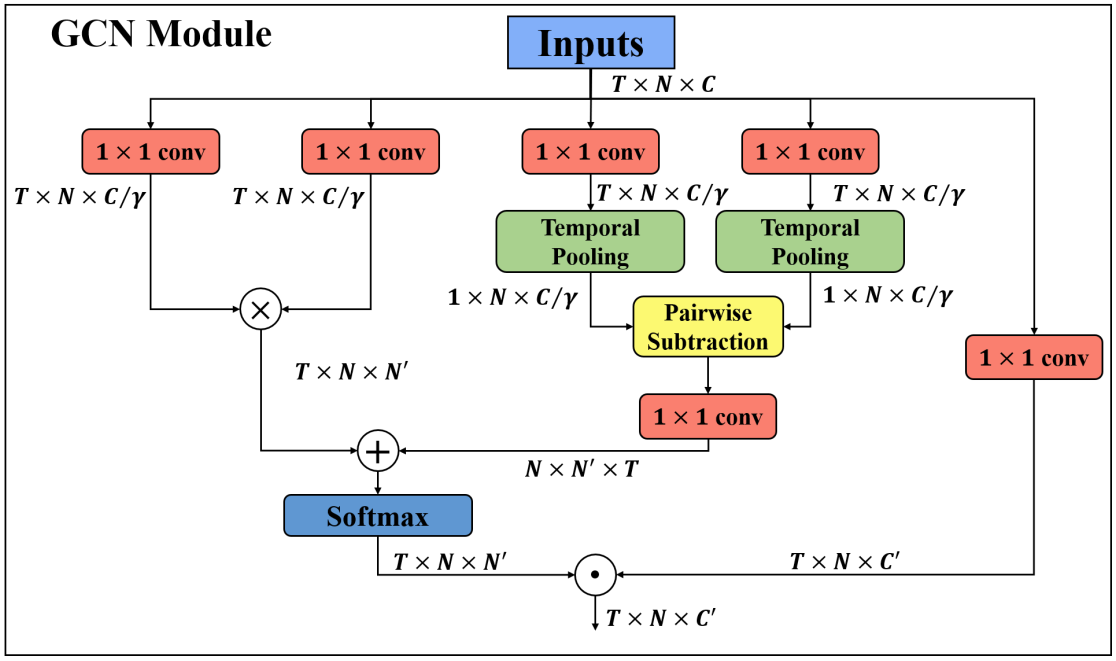


Figure 5.2: **The architecture of the GCN module. The input feature map is used as the input signal with dimension $T \times N \times C$, where T, N, and C are the number of frames, electrodes, and channels, respectively. We set the reduction rate γ to 8 in our work to extract compact representations. $\otimes$ denotes matrix multiplication operation, $\oplus$ denotes the elementwise summation, and $\odot$ denotes element-wise multiplication.**

long-range transformed feature $f(a, b)$ characterizes only the long-range spatiotemporal relationship of the electrode pair, we use the following form to define the relationship between shared bias on the channel dimension:

$$\mathbb{B}(a, b) = \delta\left(\alpha\left(\mathbb{TP}\left(X_a''\right)\right) - \beta\left(\mathbb{TP}\left(X_b''\right)\right)\right) \tag{5.5}$$

therein, the function of temporal pooling $\mathbb{TP}$ is to aggregate temporal features, whereas in our work we use mean pooling. The $\delta \in \mathbb{R}^{T \times C/8}$, $\alpha \in \mathbb{R}^{C \times C/8}$ and $\beta \in \mathbb{R}^{C \times C/8}$ are three linear embedding functions implemented by the 1×1 convolutional layer. The distances along the channel dimension $\mathbb{B}(\cdots) \in \mathbb{R}^{N \times N \times T}$ uses the nonlinear transformations to model the topological relationship on the channels. Furthermore, we use the bias for attention score calculation to update the weighting information. We update the weights using an overall attention score that is the sum of the two component weights, thus the updated weights can be formulated as follows:

$$Output_G = X^{''} \odot (\sigma (f(a,b) + \mathbb{B}(a,b))) \tag{5.6}$$

where $\odot$ is the element-wise multiplication. σ is the softmax activation function. $X^{''}$ and $Output_G$ denote input and output feature maps.

5.2.5 Temporal Modeling Module

The duration of the different sleep stages varies. Therefore, temporal modeling is also essential for sleep staging. Current methods [107, 240] still use temporal convolutions with a single fixed scale to perform temporal modeling. The feature information obtained from distant frames is very limited, and the long-range temporal dependence is not well captured, which affects the accuracy of sleep stages. It is not optimal to use temporal convolutions with a fixed kernel size to deal with the problem of sleep staging. Consequently, the multi-scale temporal features extracted by convolution kernels with different scales are fused to better model the temporal topological features. The difference from the previous method is that we use four parallel temporal convolution branches to achieve temporal modeling, as shown in Fig. 5.3. In each branch, we introduce a bottleneck architecture [192] that uses 1×1 convolution to reduce the computational cost and thus speed up the training and model inference. In addition, the first three branches of the model utilize temporal convolutions with a kernel size of 1×3, employing different dilations [193] to analyze short-term and long-term temporal dependencies, thus obtaining multi-scale temporal receptive fields. In the final branch, a 3×1 max-pooling layer is utilized to extract the most important features. Finally, we use a concatenation strategy to fuse the features. In conclusion, the temporal modeling module is proposed to extract richer temporal features from the physiological signal sequences, which can be used to capture the temporal dependencies between sleep stages and distinguish the different duration dynamics.

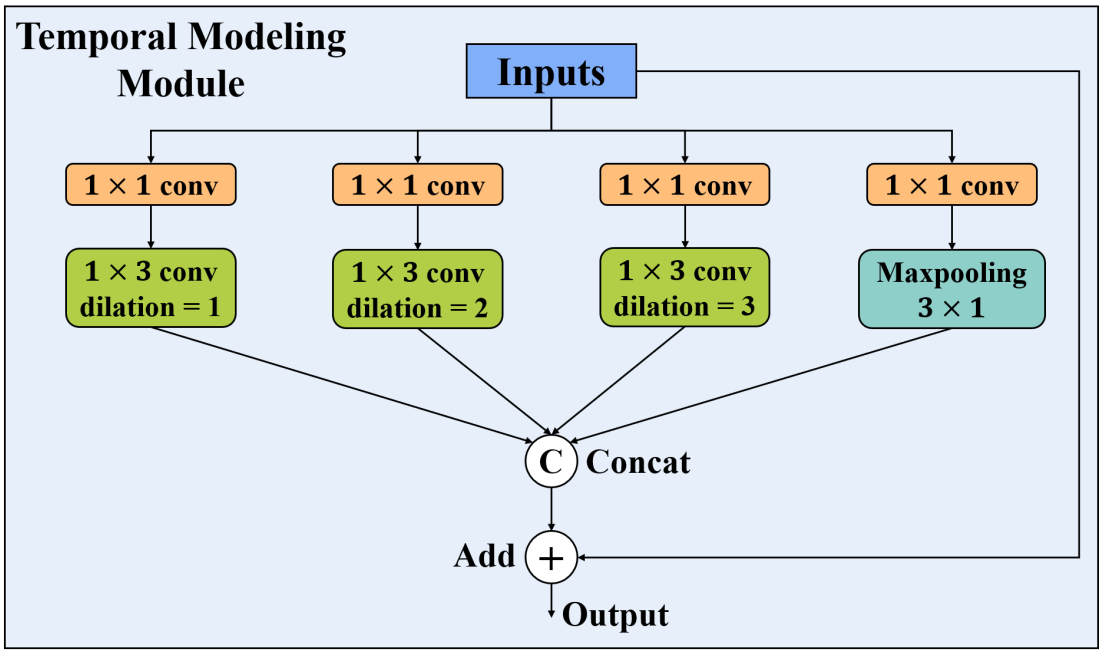


Figure 5.3: **The architecture of temporal modeling module. In order to lower the computational costs due to the extra branches, we fix kernel sizes at 1×3 and use different dilation rates for larger receptive fields. Meanwhile, the 3×1 max-pooling layer is used to capture the most salient feature.**

5.2.6 Multi-stream Fusion

In this work, we utilize multi-stream fusion strategies to model the first-order information (EEG and EOG) and the corresponding motion information for sleep stage classification. In the Sleep-EDF dataset, the sequence of electrode motion information can be obtained by calculating the difference of the same electrode between two consecutive frames, typically in terms of the differences in the coordinates of EEG and EOG electrodes. The position of the electrode of the human brain can be defined as $V_{g,t}\,\{g \in N, t \in T\}$, where N and T denote the number of electrodes and the number of frames of the signal sequences, respectively. The g represents the electrode in the frame t. As for the motion information, the position difference of the same electrode in two consecutive frames can be calculated to obtain a sequence of electrode motion information, namely the displacement information. This displacement information can then be used as an additional input feature to help the model learn dynamic features. The sequence of electrode motion information $\mathbb{M}$ for electrode g in frame t is obtained by subtracting the position of the electrode in the next frame $t+1$ from its position in

the current frame t, which can be expressed as follows:

$$\mathbb{M} = V_{g,t+1} - V_{g,t} \tag{5.7}$$

Therein, $V_{g,t+1}$ is the position of the electrode g in frame $t+1$. $\mathbb{M}$ is a vector representing the motion of the electrode between the two frames. Finally, the EEG, EOG, and corresponding motion information are fed into four streams and fused to classify different sleep stages.

5.3 Experimental Results

In this section, the effectiveness of the proposed approach is evaluated using two publicly available datasets. The first subsection provides a comprehensive description of the Sleep-EDF-39 and Sleep-EDF-153 datasets, along with the experimental setups employed in this study. Subsequently, the metrics utilized to evaluate the performance of the sleep stage model are explained. Finally, we present the performance results of our proposed model and discuss its effectiveness in comparison to other state-of-the-art models.

5.3.1 Dataset and Experimental Settings

5.3.1.1 Sleep-EDF-39 and Sleep-EDF-153 Datasets

The Sleep-EDF-39 and Sleep-EDF-153 datasets are two versions of the Sleep-EDF dataset [200]. The Sleep-EDF-153 dataset is an expanded version of the Sleep-EDF-39 dataset. The two publicly available datasets are commonly utilized in sleep staging research and are sourced from the PhysioBank. The participants are enrolled in the Sleep Cassette (SC) and Sleep Telemetry (ST) studies. In our experiment, we adopt the PSG sleep recordings from SC. They record the PSGs of healthy Caucasians without any sleep-related medications. Each subject records PSG recordings during two subsequent

day-night periods, which include two scalp-EEG, horizontal EOG, chin EMG, and event markers. Therein, EEG is sampled from F_{pz}–C_z and P_z–O_z electrde locations. In our experiments, the F_{pz}–C_z EEG is used as the input EEG signal. All EEG and EOG are acquired at a sampling rate of 100 Hz. The sleep-EDF-39 dataset contains data files for 20 male and female subjects (age 28.7 $\pm$ 2.9). The number of participants in the Sleep-EDF-153 data set is 78, ranging in age from 25 to 101 years. Consistent with some baseline approaches [182, 231], the Sleep-EDF-39 and Sleep-EDF-153 datasets in our experiment contain 41950 and 195479 sleep epochs, respectively, as shown in Table 5.2. Moreover, in two datasets, each 30-s recording is manually classified into eight stages ($wake$, S_1, S_2, S_3, S_4, REM, movement, and unknown) according to the R&K standard [1]. In the latest AASM manual [2], movement and unknown stages are excluded and the S_3 and S_4 stages are combined into one signal stage N_3. Therefore, sleep stages in the datasets consist of W (Wake), N_1 (S_1), N_2 (S_2), N_3 (S_3 + S_4) and R (REM).

Table 5.2: Details of the number of sleep stages in the sleep-EDF-39 and sleep-EDF-153 datasets.

Dataset	W	N_1	N_2	N_3	R	**Total**
Sleep-EDF-39	7927	2804	17799	5703	7717	41950
Sleep-EDF-153	65951	21522	69132	13039	25835	195479

5.3.1.2 Experimental Setting

In our experiment, the proposed model is implemented on the Pytorch platform with an RTX 3060 GPU card. The network is trained with a batch size of 64. The Adam optimizer as an optimization strategy is used to train the model for 120 epochs. The learning rate is set to 10^{-3} and is decayed by 10 at the 30^{th}, 60^{th}, and 90^{th} epochs, respectively. In our work, we set the weights of the EEG stream, the EOG stream, and the corresponding motion stream to 0.6, 0.6, 0.4, and 0.4 for weighted fusion like other multi-stream GCN methods. To improve the generalization performance and reliability

of our proposed model and reduce the risk of overfitting, we implement dropout and label smoothing [241] during the training process. Specifically, in our experimental setup, we set the dropout rate to 0.2 and employ label smoothing for better-calibrated classification networks with a smoothing factor of 0.1. In addition, we use K-fold cross-validation to evaluate the performance of our sleep staging model. We follow a rigorous evaluation methodology, using a 20-fold cross-validation scheme with K set at 20 to ensure a fair comparison with baseline models. For this purpose, subjects in the sleep-EDF-39 and sleep-EDF-153 datasets are divided into 20 groups. Accordingly, experimental results for 20-fold cross-validation are obtained. Eventually, we calculate the average of the results of all 20 test samples as the final experimental results of our model, which provide reliable performance metrics for assessing the performance of the network. Moreover, we use the TensorBoard to monitor the training progress to evaluate the performance of our proposed model on two public datasets. As shown in Figure 5.4, we observe that the training loss gradually decreases and stabilizes over iterations. This trend indicates that our model is effectively learning patterns and features from the training data.

5.3.2 Evaluation Metrics

To provide a comprehensive evaluation of the performance of the sleep staging model, we introduce several metrics including accuracy, macro-precision, macro-recall, macro-averaged F1 score, and Cohen's kappa coefficient. The overall accuracy (ACC), macro-precision (P_{macro}), macro-recall (R_{macro}), macro-averaged F1 score ($MF1$), and Cohen's kappa coefficient (κ) are defined as follows:

$$ACC = \frac{1}{K}\sum_{i=1}^{K}\left(\frac{TP+TN}{TP+FP+FN+TN}\right)_i \tag{5.8}$$

$$P_{macro} = \frac{1}{K}\sum_{i=1}^{K}\left(\frac{TP}{TP+FP}\right)_i \tag{5.9}$$

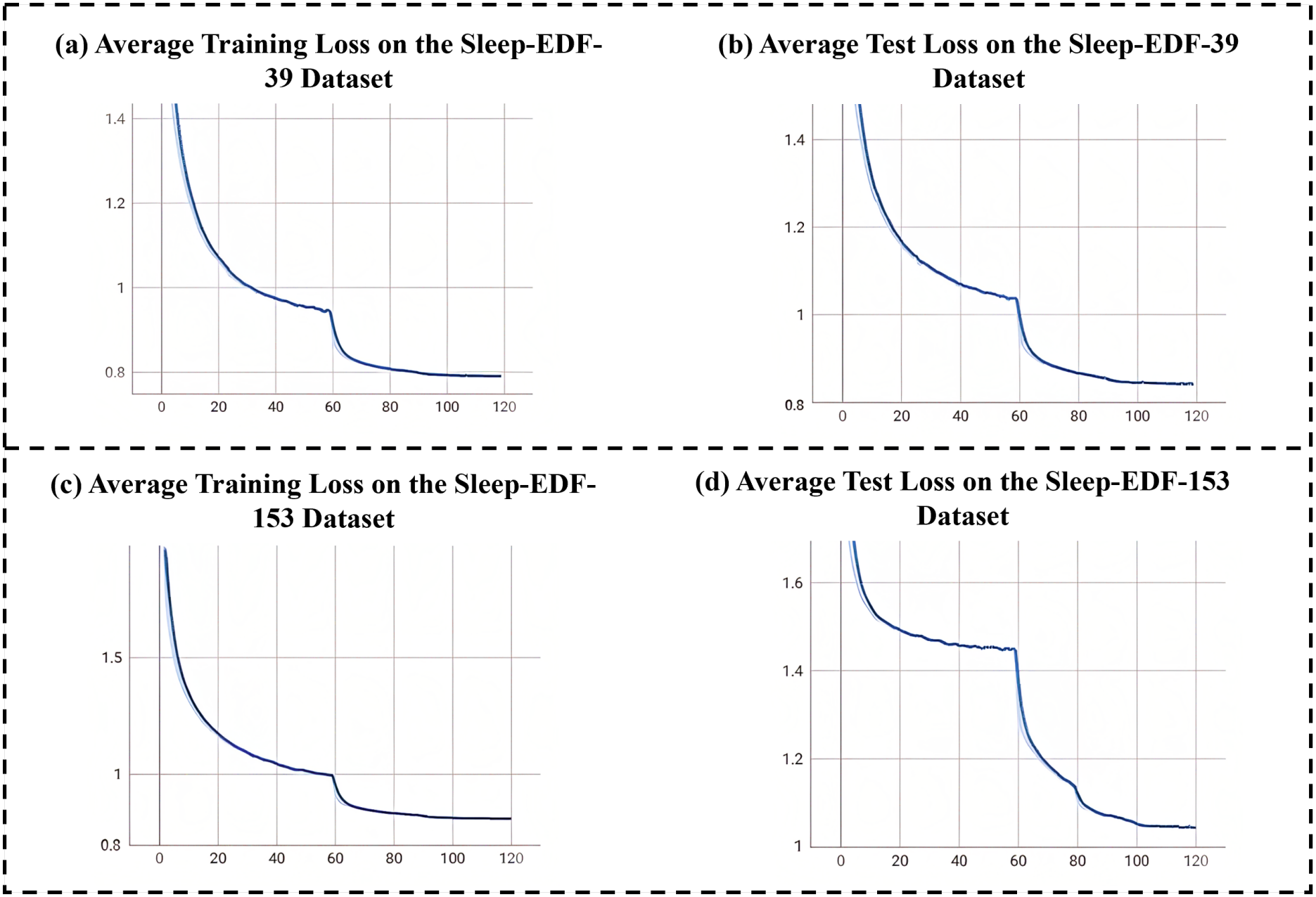


Figure 5.4: **Training and test loss vs. a number of epochs of the proposed model. The horizontal axes and the vertical axes represent epochs and the value of the loss function, respectively. The sub-figure(a) and sub-figure(b) show the training loss and test loss on the Sleep-EDF-39 dataset. The sub-figure(c) and sub-figure(d) show the proposed model loss for training and testing on the Sleep-EDF-153 dataset.**

$$R_{macro} = \frac{1}{K}\sum_{i=1}^{K}\left(\frac{TP}{TP+FN}\right)_i \tag{5.10}$$

$$MF1 = \frac{1}{K}\sum_{i=1}^{K}\left(\frac{2\times TP}{2\times TP+FN+FP}\right)_i \tag{5.11}$$

$$\kappa = \frac{ACC - p_e}{1 - p_e} \tag{5.12}$$

where TP, FP, FN, and TN stand for the true positives, false positives, false negatives, and true negatives, respectively. K represents the total number of epochs used in the cross-validation, which is defined as 20 in this work. p_e denotes the hypothetical probability of chance agreement.

5.3.3 Experiment Results

In this subsection, the effectiveness of the proposed model is evaluated using the Sleep-EDF-39 and Sleep-EDF-153 datasets. In Fig. 5.5, the confusion matrices for the predicted sleep stage of each dataset are visualized, showing agreement with the expert results. Based on Equation 5.7 and the confusion matrices, the overall accuracy of our model for the two datasets can be determined by calculation and is equal to 92.3% and 85.5%, respectively. For the Sleep-EDF-39 dataset, the macro-precision, macro-recall, and macro-F score are 88.7%, 90.0%, and 89.1%, respectively. Similarly, from the subfigure(b) of Figure 5.5, we obtain the macro-precision, macro-recall, and macro-F score of the Sleep-EDF-153 dataset as 81.9%, 80.4%, and 80.6%, respectively. Furthermore, we use Cohen's kappa coefficients to measure the degree of accuracy and reliability in sleep stage classification. The Cohen's kappa coefficients for Sleep-EDF-39 and Sleep-EDF-153 are 0.89 and 0.80, respectively, indicating that the classification results have high consistency with the actual distribution of sleep stages, being within the standard of $0.8 \sim 1$ [203].

Moreover, to investigate the effects of the classification accuracy of different sleep stages from two publicly available datasets, the receiver operating characteristic (ROC)

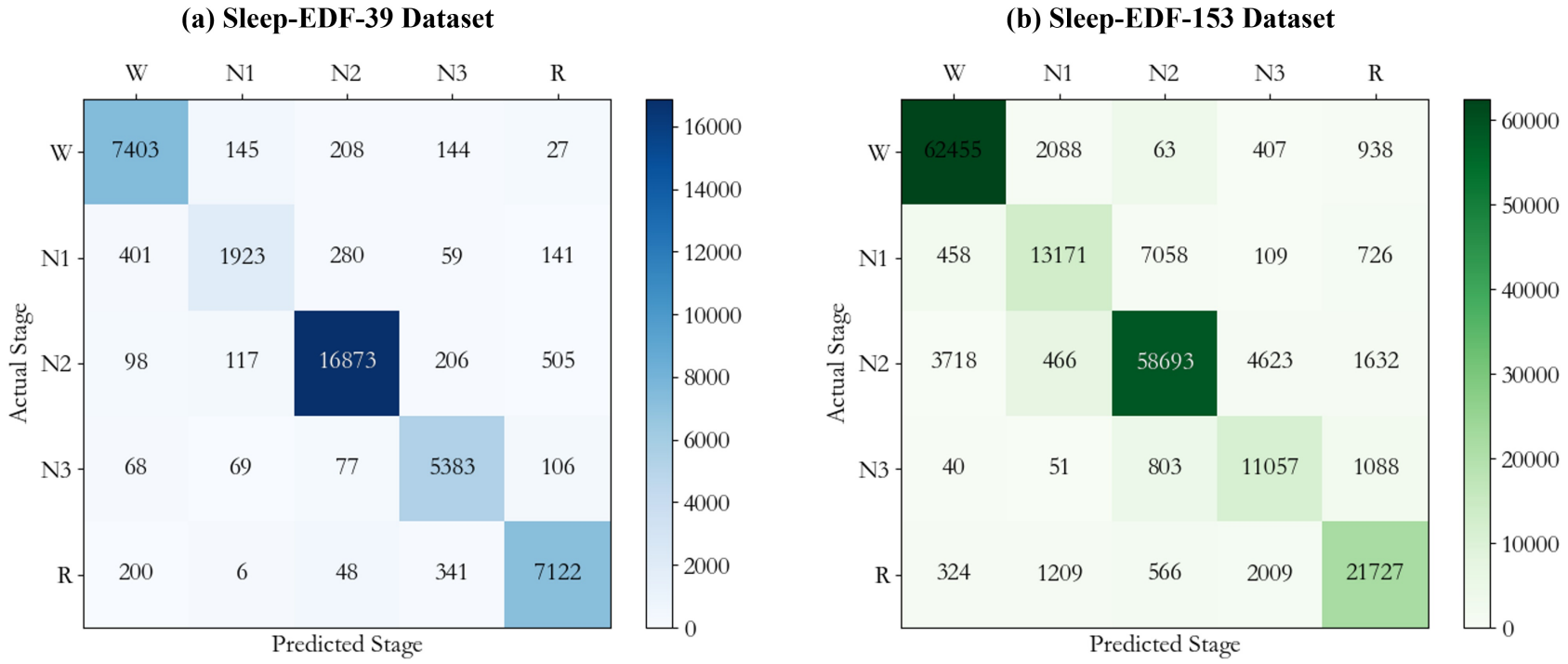


Figure 5.5: **Visualization of the experimental confusion matrix obtained from 20-fold validation. We employ the Sleep-EDF-39 and Sleep-EDF-153 datasets to obtain two confusion matrices. The sub-figure(a) and sub-figure(b) show the confusion matrix for the Sleep-EDF-39 dataset and the Sleep-EDF-153 dataset, respectively.**

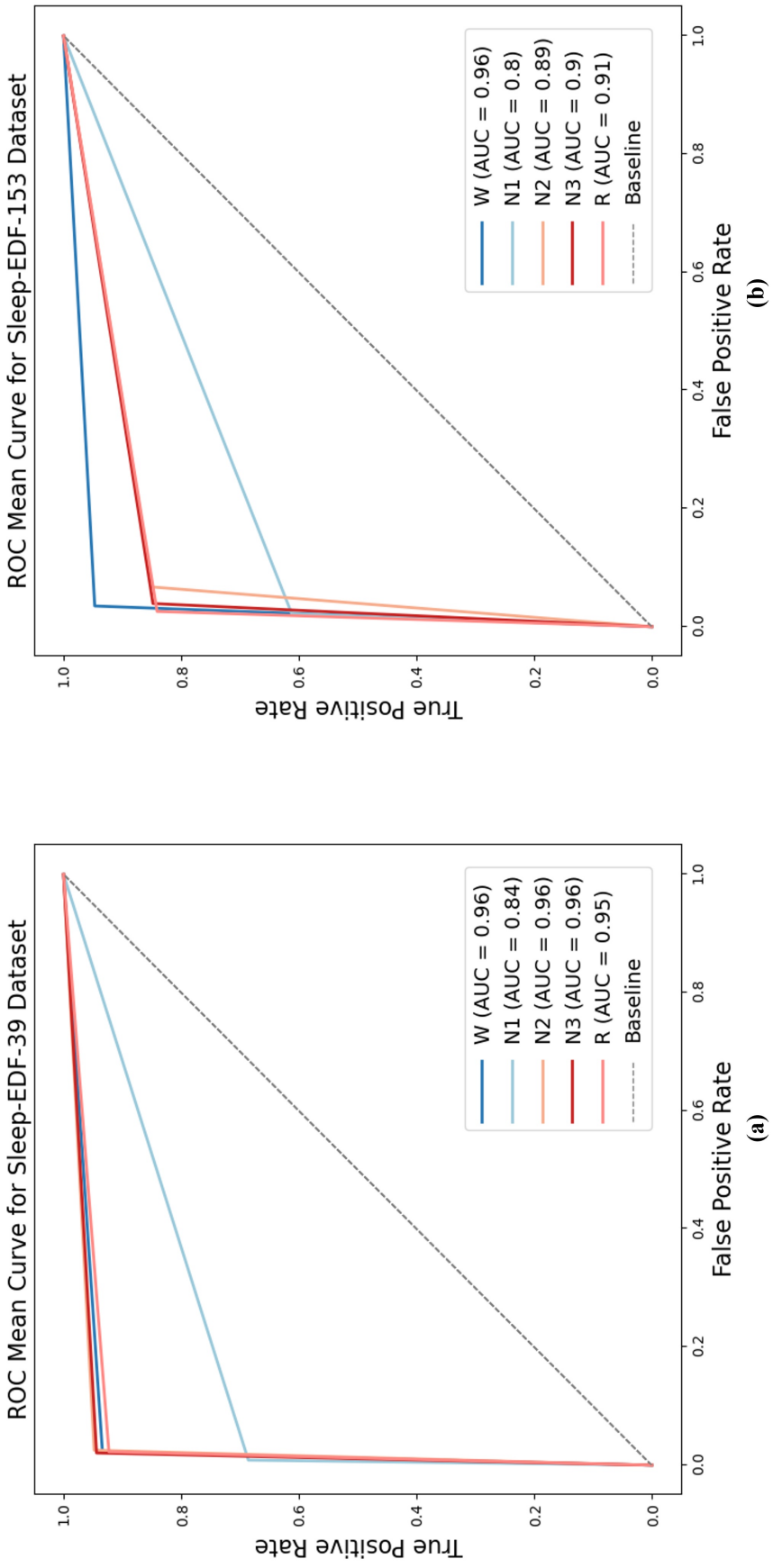


Figure 5.6: **The mean ROC curve and AUC values for different sleep stages based on 20-fold cross-validation. The ROC mean curves in sub-figure(a) and sub-figure(b) respectively use the Sleep-EDF-39 and Sleep-EDF-153 datasets as the testing datasets. The AUC values for the five sleep stages are included in the legend.**

mean curves of different sleep stages are obtained to show the effect of the proposed sleep staging model on the final classification accuracy, as shown in Figure 5.6. As expected, the ROC curves of all sleep stages, except for the N_1 stage, converge towards the upper-left corner of the graph. This convergence signifies that our model exhibits high true positive rates (TPR) and low false positive rates (FPR). This trend further demonstrates the excellent predictive performance of our model in accurately classifying different sleep stages. Nevertheless, the area under curve (AUC) values for each sleep stage (ranging from 0.8 to 0.96) on both datasets significantly exceed the value of 0.75 in [242]. This substantial improvement in AUC values underscores the superior performance of our model, which holds high clinical value. These results indicate that our proposed model not only outperforms random classification but also demonstrates a noteworthy ability to differentiate between positive and negative instances.

To further verify the advantage of the proposed multi-stream fusion strategy in sleep stage classification, we test the performance using four data modalities: *single-stream model*, which uses either the EEG or EOG stream independently; *two-stream model*, which fuses the EEG and EOG modalities; *four-stream model*, which incorporates the EEG stream, the EOG stream, the EEG motion stream, and the EOG motion stream. Table 5.3 shows that the EOG modality performs slightly better than the EEG modality in sleep staging. The superiority of the multi-stream method over the single-stream method is evident. Compared to the two-stream model, we respectively obtain 0.8% and 1.1% improvement on two datasets with the fusion of all four streams. This suggests that the fusion of the EEG stream, the EOG stream, and the corresponding motion stream can yield better classification performance, thus becoming a better choice for sleep stage classification.

5.3.4 Comparison with State-of-the-Art Models

To evaluate the effectiveness of our proposed method, we conduct a comparison between our proposed 4s-SleepGCN model and several baseline models using the Sleep-EDF-39 and Sleep-EDF-153 datasets. The results of this comparison are presented in

Table 5.3: Comparisons of the validation results with different input modalities on Sleep-EDF-39 and Sleep-EDF-153 datasets.

Methods	Acc. I (%)	Acc. II (%)
1s-SleepGCN (only EEG)	89.2	82.1
1s-SleepGCN (only EOG)	89.8	82.8
2s-SleepGCN	91.5	84.4
4s-SleepGCN	92.3	85.5

[1] 2s-SleepGCN represents using the EEG and EOG modalities.
[2] 4s-SleepGCN represents using EEG stream, EOG stream, EEG motion stream, and EOG motion stream.
[3] Acc. I and Acc. II shows the overall accuracy for Sleep-EDF-39 and Sleep-EDF-153 datasets, respectively.

Table 5.4. In comparison to other baseline methods, our method reaches state-of-the-art accuracy of 92.3% and 85.5%, outperforming the baseline models by more than 1.3% and 0.2% on two public datasets.

For some traditional machine learning-based methods, e.g., SVM and RF, the inability to adequately extract various features often leads to poor results in sleep stage classification. Deep learning methods have become a predominant approach for sleep stage classification to achieve better performance, including those using only CNNs, only GCNs, and a mixture of CNNs and RNNs. Despite the fact that these methods perform reasonably well in the sleep stage classification, resulting in varying degrees of drawbacks. For instance, it is difficult to adjust and optimize some mixed deep-learning models with extensive parameters such as DeepSleepNet, SeqSleepNet, and TinySleepNet. Moreover, there are also methods, e.g., ResnetLSTM and SleepEEGNet, that convert physiological signals into time-frequency images, which often leads to partial information loss. This contrasts with the previous work, where our model uses a multi-information flow fusion method to capture the distinctive complementary features of the original data. Moreover, the motion information from EEG and EOG aids in further enhancing the performance of sleep stage classification. Therefore, our proposed 4s-SleepGCN achieves the highest accuracy compared with other baseline

Table 5.4: Performance of the Sleep-EDF-39 and Sleep-EDF-153 datasets compared with baseline methods.

Methods	Sleep-EDF-39 dataset							Sleep-EDF-153 dataset						
	Overall results		F1-Score for Sleep Stag(%)					Overall results		F1-Score for Sleep Stag(%)				
	Macro-F score(%)	Accuracy(%)	$Wake$	N_1	N_2	N_3	REM	Macro-F score(%)	Accuracy(%)	$Wake$	N_1	N_2	N_3	REM
SVM [231]	63.7	76.1	71.6	13.6	85.1	76.5	71.8	57.8	71.2	80.3	13.5	79.5	57.1	58.7
RF [231]	67.6	78.1	74.9	22.5	86.3	80.8	73.3	62.4	72.7	81.6	23.2	80.6	65.8	60.8
SleepEEGNet [180]	79.7	84.3	89.2	52.2	86.8	85.1	85.0	77.0	82.8	90.3	44.6	85.7	81.6	82.9
U-time [243]	78.6	78.2	87.0	52.0	86.0	84.0	84.0	76.4	-	92.0	51.0	84.0	75.0	80.0
MultitaskCNN [179]	75.0	83.1	87.9	33.5	87.5	85.8	80.3	72.8	79.6	90.9	39.7	83.2	76.6	73.5
AttnSleep [244]	78.1	84.4	89.7	42.6	88.8	90.2	79.0	75.1	81.3	92.0	42.0	85.0	82.1	74.2
DeepSleepNet [182]	76.9	82.0	84.7	46.6	85.9	84.8	82.4	75.3	78.5	91.0	47.0	81.0	69.0	79.0
TinySleepNet [183]	80.5	85.4	90.1	51.4	88.5	88.3	84.3	78.1	83.1	92.8	51.0	85.3	81.1	80.3
SeqSleepNet [175]	79.7	86.0	91.9	47.8	87.2	85.7	86.2	78.2	83.8	92.8	48.9	85.4	78.6	85.1
ResnetLSTM [245]	73.7	82.5	86.5	28.4	87.7	89.8	76.2	71.4	78.9	90.7	34.7	83.6	80.9	67.0
MLTCN [246]	77.1	84.2	88.5	39.4	87.7	87.0	82.7	74.9	81.0	92.2	42.8	83.3	**88.3**	77.7
SleepPrintNet [232]	78.0	83.1	88.8	48.0	86.7	86.2	80.3	76.5	81.6	92.7	47.4	83.6	80.0	78.8
SalientSleepNet [231]	83.0	87.5	92.3	56.2	89.9	87.2	89.2	79.5	84.1	93.3	54.2	85.8	78.3	**85.8**
Our first work [119]	89.0	91.0	**92.1**	**79.7**	93.2	88.2	**91.6**	**81.1**	85.3	92.9	66.6	86.0	75.2	84.6
4s-SleepGCN (ours)	**89.1**	**92.3**	92.0	75.9	**95.6**	**91.0**	91.2	80.6	**85.5**	**94.0**	**68.4**	**86.1**	70.7	83.7

The numbers in bold indicate the highest performance metrics among all approaches, while the result underlined represents the sub-optimal performance.

models.

On the Sleep-EDF-153 dataset, the classification performance of the W and N_2 stages is the best among all sleep stages. Specifically, the F1 score of the W and N_2 stages reaches 94.0% and 86.1%, respectively. Moreover, for this reason, the N_1 stage belongs to the sleep transition period [205], which can be mainly misclassified into N_2 and REM stages. The classification effect for the N_1 stage falls short of expectations compared to the other sleep stages, but it still achieves an optimal result compared to the other baseline methods. This is sufficient to illustrate that our model can effectively classify sleep stages in a large sample dataset. Additionally, we can observe that the F1 score of N_3 and REM stages is worse than that of most baseline models. The poor results attributed to the fact that $N_3 - N_2$ and $REM - N_2$ are also misclassified pairs. In classifying the N_3 stage, an important factor contributing to its lower classification performance is the small proportion of N_3 stage instances within the Sleep-EDF-153 dataset, representing only 6.67% of the total. The limited number of N_3 stage examples in the dataset poses a challenge for the classification model to effectively learn the specific patterns and features associated with the N_3 stage. Due to this scarcity, our proposed model may not be sufficiently familiar with the minority class, resulting in suboptimal generalization and a drop in performance in classifying the N_3 stage. However, the precision of the N_3 and REM stages reaches 84.8% and 84.0%, respectively. Therefore, our proposed model can to a large extent reproduce the sleep scoring of human experts and thus provide assistance in the diagnosis of sleep problems.

Besides, we show the comparative results in terms of accuracy and model complexity (number of parameters) with some state-of-the-art methods to demonstrate the superiority of our model. As can be seen in Table 5.5, the efficiency of our model has improved compared to previous models for the Sleep-EDF-39 dataset. At first glance, our proposed 4s-SleepGCN has a larger number of parameters than SalientSleepNet. However, our method has adopted the four-stream network architecture, which consists of four backbones. In comparison, the proposed single-stream model based on the EEG or EOG modality achieves relatively great results with an accuracy of 89.2%

Table 5.5: Comparison of model parameters on Sleep-EDF-39 dataset.

Methods	Param.(M)	Acc.(%)
SleepEEGNet [180]	2.1	84.3
TinySleepNet [183]	1.3	85.4
U-time [243]	1.1	78.2
SalientSleepNet [231]	0.9	87.5
1s-SleepGCN (only EEG)	0.6	89.2
1s-SleepGCN (only EOG)	0.6	89.8
2s-SleepGCN	1.2	91.5
4s-SleepGCN	2.5	92.3

[1] The Acc. denotes the accuracy for Sleep-EDF-39 dataset.
[2] 2s-SleepGCN represents using the EEG and EOG modalities.
[3] 4s-SleepGCN represents using EEG stream, EOG stream, EEG motion stream, and EOG motion stream.

and 89.8%, respectively. Besides, the proposed single-stream model requires only 0.6 million parameters, which reduces the number of parameters by about 0.3 million. This proves that our proposed single-stream solid baseline can be introduced as a strong and powerful baseline for sleep stage classification. The proposed 2s-SleepGCN and 4s-SleepGCN require about 0.3M+ and 1.6M+ more parameters compared to the SalientSleepNet, while improving the accuracy by 4% and 4.8%, respectively. We conclude that the lightweight, single-stream solid baseline constructed in this study can significantly reduce the number of model parameters while ensuring classification accuracy. In addition, the two-stream and four-stream proposals show better performance when more parameters are requested.

5.4 Discussion

Sleep disorders have indeed risen in striking proportion worldwide over the past 40 years [208, 247, 248]. Sleep stage classification plays a critical role in the diagnosis and treatment of sleep disorders. Automated sleep stage scoring is expected to play a leading role in the diagnosis and treatment of sleep disorders in the future. In this work, a

graph-based multi-stream fusion model named 4s-SleepGCN is proposed for sleep stage classification. EEG, EOG, and the corresponding motion information are fused to enhance the understanding of brain activity and aid in the identification of different sleep stages. This confirms that the motion modality holds significant potential for sleep stage classification and contributes to improved accuracy and temporal understanding of sleep stages. The proposed EEG or EOG single-stream method with a lightweight network has demonstrated acceptable performance on benchmark datasets, making it a promising candidate for application in residential healthcare settings. In clinical medicine, there is a need to accurately classify different sleep stages and provide reliable results for specialists. The proposed multi-stream model holds the potential to assist doctors in making accurate diagnostic and treatment decisions, thereby improving patients' sleep health outcomes.

The Sleep-EDF-39 dataset and Sleep-EDF-153 dataset utilize in our study comprise practical data obtained from patients. It is important to note that these datasets are non-independent and non-identically distributed, meaning there are significant variations in the sample sizes across different sleep stages. Nevertheless, our proposed method demonstrates robustness by achieving satisfactory classification results for each sleep stage. This also underscores its effectiveness in handling the complexities inherent in real-world patient data. In addition, our proposed multi-stream model demonstrates remarkable classification performance, particularly in the N_2 stage. Abnormalities observed in N_2 sleep features have been identified as potential indicators for various sleep disorders such as sleep apnea and parasomnias. The accurate classification of the N_2 stage by our model holds significant promise in the identification, diagnosis, and intervention of sleep disorders, ultimately leading to enhanced sleep quality and overall well-being. The exceptional classification performance of our multi-stream model, particularly in the N_2 stage, highlights its potential as a valuable tool in sleep research, clinical assessments, and interventions aimed at optimizing sleep architecture. Its robust capabilities make it an asset in furthering our understanding of sleep-related phenomena and facilitating effective interventions to address sleep disorders. By leveraging the

strengths of our proposed model, researchers and clinicians can make significant strides in the field of sleep medicine, ultimately improving the lives of individuals affected by sleep-related issues. Furthermore, for the Sleep-EDF-153 and Sleep-EDF-39 datasets, the ratio of the average training time per fold (approximately 4.17 and 1.36 hours, respectively) is smaller than the ratio of the respective data sizes (195k and 42k). In other words, the training time of our proposed model does not increase proportionally to the increase in data size. Therefore, our model can effectively manage the processing of larger datasets without significantly increasing the training time. This indicates that the proposed model demonstrates a certain degree of scalability. Such scalability is particularly valuable in real-world scenarios where the volume of data is substantial.

Chapter 6

Conclusion and Future Work

In our first work, we propose a combination of dynamic and static ST-GCN with inter-temporal attention blocks for automatic sleep stage classification based on EEG. Spatial graph convolutions and temporal convolutions are used to model the EEG data. We use a combination of dynamic and static ST-GCN to capture the global context-enriched topology and employ temporal convolution with dilation to enlarge the temporal receptive field. Furthermore, to the best of our knowledge, we introduce the attention blocks for the first time in the field of sleep stage classification to model the relationship between different EEG channels, which can capture long-range dependencies for sleep stage classification. The comparative results indicate that our method has powerful capability and expressiveness in sleep stage classification. Therefore, we believe that our method could be a complementary tool to help scientists monitor the sleep status of patients to initiate appropriate treatments. In the future, since our method is used for sleep stage classification based on EEGs, we will apply it to a broader range of other physiological signal classification tasks.

Moreover, in our second proposed work, we propose a novel multi-stream fusion graph convolutional network called 4s-SleepGCN to efficiently classify different sleep stages by combining multi-stream biological signal features. The positional relationship of modal sequences is embedded into the sleep staging network to improve the feature characterization capability, which can better leverage the task of sleep stage

classification. Besides, the proposed 4s-SleepGCN model uses graph convolution and temporal convolution to directly model spatial-temporal dependencies from the PSG graph sequences. Graph convolution can effectively extract the long-range dependencies between electrodes. Temporal convolution can learn richer temporal features and aggregate multi-scale contextual information. Furthermore, we model EEG, EOG, and the corresponding motion information in a unified multi-stream network framework for the first time, demonstrating the validity of motion modality. Experiments on the Sleep-EDF-39 and Sleep-EDF-153 datasets evaluate the feasibility and superiority of our proposed model. Our proposed 4s-SleepGCN model achieves significantly better accuracy on both of them than the current state-of-the-art model. In addition, the proposed lightweight single-stream network with only 0.6 million model parameters achieves higher accuracy and smaller network size compared to some baseline models, which provides a new perspective in the field of sleep staging and thus can be used to monitor and track sleep in a home environment. The proposed multi-stream model can be used as a powerful tool to assist sleep experts in assessing sleep quality and diagnosing sleep-related diseases. The flexibility and adaptability of our proposed model make it suitable for various applications beyond sleep stage classification, such as medical applications, healthcare monitoring, and sports analysis.

GCN is a type of neural network architecture used in various applications, including sleep staging. Sleep staging is the process of classifying different stages of sleep based on electroencephalogram (EEG) and other physiological signals. GCNs can be applied to sleep staging by modeling the relationships between different EEG channels or other signals in a more sophisticated way compared to traditional machine learning approaches. The field of sleep staging is continually evolving, and there are several areas of future work that we are likely to focus on to improve the accuracy, efficiency, and applicability of sleep staging techniques. Here are some potential directions for future research:

1. In our work, EEG, EOG, and the corresponding motion information are fused for sleep stage classification and our proposed approach significantly improves the

accuracy and depth of sleep stage classification. In addition to EEG and EOG, we can explore the integration of other physiological signals such as EMG, heart rate variability (HRV), respiratory rate, and blood oxygen levels (SpO2). Combining these signals can provide a more comprehensive view of a person's sleep. Future work may involve developing models that can effectively integrate and analyze data from these various modalities to enhance the accuracy of sleep stage classification.

2. Classifying sleep stages, especially the N_1 stage, can remain challenging due to its transitional nature between wakefulness and sleep, making correct recognition a tricky task. In the future, we can use a method of combining clinical validation and expert feedback. Collaborations between sleep experts and deep learning researchers can lead to better-defined criteria for stage N_1 classification. Ongoing clinical validation studies can help refine algorithms and ensure their accuracy in real-world settings. Improving the accuracy of Stage N1 detection is important not only for understanding sleep dynamics but also for diagnosing sleep disorders and providing targeted interventions. As technology advances and our understanding of sleep physiology deepens, we can expect ongoing progress in this area.

3. Sleep stage classification typically relies on the subjective interpretation and classification of physiological signals by experts. Nonetheless, different experts may interpret the same data and arrive at varying conclusions. Namely, It is inevitable that similar sleep stages may be incorrectly marked. Therefore, the question for many sleep stage classification networks is how to use high-quality sleep stage datasets for the training process. Implementing validation procedures and quality control measures in sleep laboratories can help monitor and improve the consistency of expert scoring. Regular audits and checks can identify and rectify potential sources of variability. In addition, we will continue to explore this area and leverage advanced technologies to develop a sleep stage classification system that provides a more human-like performance classification model. This can

not only assist experts by providing additional information and insight during the scoring process. But also it can flag potential discrepancies for further review.

4. Existing sleep staging models are typically processed offline, analyzing and capturing post-sleep data. However, for the timely detection and intervention of potential sleep issues, real-time monitoring is crucial. Real-time monitoring systems can analyze an individual's sleep data and provide personalized sleep recommendations. These recommendations may include adjustments to bedtime routines, sleep environment, and lifestyle factors based on the user's specific sleep patterns and goals. Moreover, real-time monitoring can aid in the early detection of sleep disorders. Algorithms can analyze continuous sleep data to identify patterns indicative of conditions such as sleep apnea, insomnia, or restless leg syndrome, enabling timely intervention and treatment. This is particularly significant for patients with sleep apnea, as real-time detection enables the adjustment of ventilation pressure and treatment parameters, leading to optimized treatment outcomes. Therefore, the future of real-time monitoring and feedback in sleep staging involves leveraging technology to empower individuals to take an active role in managing their sleep health. These advancements can lead to more personalized and effective sleep solutions, early detection of sleep disorders, and improved overall well-being.

References

[1] N. A. Asif, Y. Sarker, R. K. Chakrabortty, M. J. Ryan, M. H. Ahamed, D. K. Saha, F. R. Badal, S. K. Das, M. F. Ali, S. I. Moyeen *et al.*, "Graph neural network: A comprehensive review on non-euclidean space," *IEEE Access*, vol. 9, pp. 60 588–60 606, 2021, doi:10.1109/ACCESS.2021.3071274.

[2] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[3] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: a comprehensive review," *Computational Social Networks*, vol. 6, no. 1, pp. 1–23, 2019, doi:10.1186/s40649-019-0069-y.

[4] S. Fatma and R. Anjum, "Sleep: Physiology & its concept in unani medicine," *International Journal of Advance Research and Innovative Ideas in Education*, vol. 6, pp. 1321–1327, 2020.

[5] N. S. Murali, A. Svatikova, and V. K. Somers, "Cardiovascular physiology and sleep," *Frontiers in Bioscience*, vol. 8, no. 6, pp. s636–52, 2003, doi:10.2741/1105.

[6] E. Ezenwanne, "Current concepts in the neurophysiologic basis of sleep; a review," *Annals of medical and health sciences research*, vol. 1, no. 2, pp. 173–180, 2011.

[7] F. S. Luyster, P. J. Strollo Jr, P. C. Zee, and J. K. Walsh, "Sleep: a health imperative," *Sleep*, vol. 35, no. 6, pp. 727–734, 2012, doi:10.5665/sleep.1846.

[8] J. A. Groeger, F. Zijlstra, and D.-J. Dijk, "Sleep quantity, sleep difficulties and their perceived consequences in a representative sample of some 2000 british adults," *Journal of sleep research*, vol. 13, no. 4, pp. 359–371, 2004, doi:10.1111/j.1365-2869.2004.00418.x.

[9] T. Paunio, T. Korhonen, C. Hublin, M. Partinen, M. Kivimäki, M. Koskenvuo, and J. Kaprio, "Longitudinal study on poor sleep and life dissatisfaction in a nationwide cohort of twins," *American Journal of Epidemiology*, vol. 169, no. 2, pp. 206–213, 2009, doi:10.1093/aje/kwn305.

[10] T. Ohara, T. Honda, J. Hata, D. Yoshida, N. Mukai, Y. Hirakawa, M. Shibata, H. Kishimoto, T. Kitazono, S. Kanba *et al.*, "Association between daily sleep duration and risk of dementia and mortality in a japanese community," *Journal of the American Geriatrics Society*, vol. 66, no. 10, pp. 1911–1918, 2018, doi:10.1111/jgs.15446.

[11] R. Robbins, S. F. Quan, M. D. Weaver, G. Bormes, L. K. Barger, and C. A. Czeisler, "Examining sleep deficiency and disturbance and their risk for incident dementia and all-cause mortality in older adults across 5 years in the united states," *Aging (Albany NY)*, vol. 13, no. 3, p. 3254, 2021, doi:10.18632/aging.202591.

[12] M. He, X. Deng, Y. Zhu, L. Huan, and W. Niu, "The relationship between sleep duration and all-cause mortality in the older people: an updated and dose-response meta-analysis," *BMC Public Health*, vol. 20, pp. 1–18, 2020, doi:10.1186/s12889-020-09275-3.

[13] E. Bixler, "Sleep and society: an epidemiological perspective," *Sleep medicine*, vol. 10, pp. S3–S6, 2009, doi:10.1016/j.sleep.2009.07.005.

[14] K. Kim, M. Uchiyama, M. Okawa, X. Liu, and R. Ogihara, "An epidemiological study of insomnia among the japanese general population," *Sleep*, vol. 23, no. 1, pp. 41–47, 2000.

[15] K. Suzuki, M. Miyamoto, and K. Hirata, "Sleep disorders in the elderly: Diagnosis and management," *Journal of general and family medicine*, vol. 18, no. 2, pp. 61–71, 2017, doi:10.1002/jgf2.27.

[16] T. M. Buckley and A. F. Schatzberg, "On the interactions of the hypothalamic-pituitary-adrenal (HPA) axis and sleep: normal HPA axis activity and circadian rhythm, exemplary sleep disorders," *The Journal of Clinical Endocrinology & Metabolism*, vol. 90, no. 5, pp. 3106–3114, 2005, doi:10.1210/jc.2004-1056.

[17] I. Pollicina, A. Maniaci, J. R. Lechien, G. Iannella, C. Vicini, G. Cammaroto, A. Cannavicci, G. Magliulo, A. Pace, S. Cocuzza *et al.*, "Neurocognitive performance improvement after obstructive sleep apnea treatment: state of the art," *Behavioral Sciences*, vol. 11, no. 12, p. 180, 2021, doi:10.3390/bs11120180.

[18] A. N. Garcia and I. M. Salloum, "Polysomnographic sleep disturbances in nicotine, caffeine, alcohol, cocaine, opioid, and cannabis use: a focused review," *The American journal on addictions*, vol. 24, no. 7, pp. 590–598, 2015, doi:10.1111/ajad.12291.

[19] S. Ram, H. Seirawan, S. K. Kumar, and G. T. Clark, "Prevalence and impact of sleep disorders and sleep habits in the united states," *Sleep and breathing*, vol. 14, pp. 63–70, 2010, doi:10.1007/s11325-009-0281-3.

[20] M. M. e Cruz, M. H. Kryger, C. M. Morin, L. Palombini, C. Salles, and D. Gozal, "Comorbid insomnia and sleep apnea: Mechanisms and implications of an underrecognized and misinterpreted sleep disorder," *Sleep Medicine*, vol. 84, pp. 283–288, 2021, doi:10.1016/j.sleep.2021.05.043.

[21] B. M. Altevogt, H. R. Colten *et al.*, "Sleep disorders and sleep deprivation: an unmet public health problem," 2006, doi:10.17226/11617.

[22] S.-F. Liang, C.-E. Kuo, Y.-H. Hu, and Y.-S. Cheng, "A rule-based automatic sleep staging method," *Journal of neuroscience methods*, vol. 205, no. 1, pp. 169–176, 2012, doi:10.1109/IEMBS.2011.6091499.

[23] S. J. Redmond and C. Heneghan, "Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 3, pp. 485–496, 2006, doi:10.1109/TBME.2005.869773.

[24] D. Alvarez, R. Hornero, J. V. Marcos, F. Del Campo, and M. Lopez, "Spectral analysis of electroencephalogram and oximetric signals in obstructive sleep apnea diagnosis," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2009, pp. 400–403, doi:10.1109/IEMBS.2009.5334905.

[25] M. J. Sateia, "International classification of sleep disorders," *Chest*, vol. 146, no. 5, pp. 1387–1394, 2014, doi:10.1378/chest.14-0970.

[26] T. Roth and T. Roehrs, "Insomnia: epidemiology, characteristics, and consequences," *Clinical cornerstone*, vol. 5, no. 3, pp. 5–15, 2003, doi:10.1016/s1098-3597(03)90031-7.

[27] P. J. Strollo Jr and R. M. Rogers, "Obstructive sleep apnea," *New England Journal of Medicine*, vol. 334, no. 2, pp. 99–104, 1996, doi:10.1056/NEJM199601113340207.

[28] T. E. Scammell, "Narcolepsy," *New England Journal of Medicine*, vol. 373, no. 27, pp. 2654–2662, 2015, doi:10.1056/NEJMra1500587.

[29] C. Guilleminault, C. Kirisoglu, A. C. da Rosa, C. Lopes, and A. Chan, "Sleepwalking, a disorder of nrem sleep instability," *Sleep medicine*, vol. 7, no. 2, pp. 163–170, 2006, doi:10.1016/j.sleep.2005.12.006.

[30] J. D. Kales, A. Kales, C. R. Soldatos, A. B. Caldwell, D. S. Charney, and E. D. Martin, "Night terrors: Clinical characteristics and personality patterns," *Archives of General Psychiatry*, vol. 37, no. 12, pp. 1413–1417, 1980, doi:10.1001/archpsyc.1980.01780250099012.

[31] C. H. Schenck and M. W. Mahowald, "REM sleep behavior disorder: clinical, developmental, and neuroscience perspectives 16 years after its formal identification in sleep," *Sleep*, vol. 25, no. 2, pp. 120–138, 2002, doi:10.1093/sleep/25.2.120.

[32] K.-C. Lan, D.-W. Chang, C.-E. Kuo, M.-Z. Wei, Y.-H. Li, F.-Z. Shaw, and S.-F. Liang, "Using off-the-shelf lossy compression for wireless home sleep staging," *Journal of neuroscience methods*, vol. 246, pp. 142–152, 2015, doi:10.1016/j.jneumeth.2015.03.013.

[33] L. Fraiwan, K. Lweesy, N. Khasawneh, M. Fraiwan, H. Wenz, and H. Dickhaus, "Time frequency analysis for automated sleep stage identification in full-term and preterm neonates," *Journal of medical systems*, vol. 35, pp. 693–702, 2011, doi:10.1007/s10916-009-9406-2.

[34] A. J. Boe, L. L. McGee Koch, M. K. O'Brien, N. Shawen, J. A. Rogers, R. L. Lieber, K. J. Reid, P. C. Zee, and A. Jayaraman, "Automating sleep stage classification using wireless, wearable sensors," *NPJ digital medicine*, vol. 2, no. 1, p. 131, 2019, doi:10.1038/s41746-019-0210-1.

[35] F. Ebrahimi, M. Mikaili, E. Estrada, and H. Nazeran, "Assessment of itakura distance as a valuable feature for computer-aided classification of sleep stages," in *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007, pp. 3300–3303, doi:10.1109/IEMBS.2007.4353035.

[36] C. A. Kushida, A. Chang, C. Gadkary, C. Guilleminault, O. Carrillo, and W. C. Dement, "Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients," *Sleep medicine*, vol. 2, no. 5, pp. 389–396, 2001, doi:10.1016/s1389-9457(00)00098-8.

[37] J. V. Rundo and R. Downey III, "Polysomnography," *Handbook of clinical neurology*, vol. 160, pp. 381–392, 2019, doi:10.1016/B978-0-444-64032-1.00025-4.

[38] R. Caton, "Electrical currents of the brain," *The Journal of Nervous and Mental Disease*, vol. 2, no. 4, p. 610, 1875.

[39] D. Millett, "Hans berger: From psychic energy to the EEG," *Perspectives in biology and medicine*, vol. 44, no. 4, pp. 522–542, 2001, doi:10.1353/pbm.2001.0070.

[40] M. Engstrøm, E. Rugland, and M. S. Heier, "Polysomnography (PSG) for studying sleep disorders," *Tidsskrift for den Norske laegeforening: tidsskrift for praktisk medicin, ny raekke*, vol. 133, no. 1, pp. 58–62, 2013, doi:10.4045/tidsskr.12.0172.

[41] S. A. Keenan, "An overview of polysomnography," *Handbook of clinical neurophysiology*, vol. 6, pp. 33–50, 2005, doi:10.1016/S1567-4231(09)70028-0.

[42] E. Estrada, H. Nazeran, J. Barragan, J. R. Burk, E. A. Lucas, and K. Behbehani, "EOG and EMG: Two important switches in automatic sleep stage classification," in *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2006, pp. 2458–2461, doi:10.1109/IEMBS.2006.260075.

[43] A. Malhotra, M. Younes, S. T. Kuna, R. Benca, C. A. Kushida, J. Walsh, A. Hanlon, B. Staley, A. I. Pack, and G. W. Pien, "Performance of an automated polysomnography scoring system versus computer-assisted manual scoring," *Sleep*, vol. 36, no. 4, pp. 573–582, 2013, doi:10.5665/sleep.2548.

[44] E. A. Wolpert, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," *Archives of General Psychiatry*, vol. 20, no. 2, pp. 246–247, 1969, doi:10.1001/archpsyc.1969.01740140118016.

[45] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. Marcus, B. V. Vaughn *et al.*, "The AASM manual for the scoring of sleep and associated events," *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine*, vol. 176, p. 2012, 2012.

[46] L. Fiorillo, A. Puiatti, M. Papandrea, P.-L. Ratti, P. Favaro, C. Roth, P. Bargiotas, C. L. Bassetti, and F. D. Faraci, "Automated sleep scoring: A review of the latest approaches," *Sleep medicine reviews*, vol. 48, p. 101204, 2019, doi:10.1016/j.smrv.2019.07.007.

[47] C. Affonso, A. L. D. Rossi, F. H. A. Vieira, A. C. P. de Leon Ferreira *et al.*, "Deep learning for biological image classification," *Expert systems with applications*, vol. 85, pp. 114–122, 2017, doi:10.1016/j.eswa.2017.05.039.

[48] M. Li, H. Chen, and Z. Cheng, "A lightweight end-to-end network for wearing mask recognition on low-resolution images," in *IEEE 15th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC)*, 2022, pp. 38–44, doi:10.1109/MCSoC57363.2022.00016.

[49] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," *arXiv preprint arXiv:1712.04621*, 2017, doi:10.48550/arXiv.1712.04621.

[50] V. Sharma, M. Gupta, A. Kumar, and D. Mishra, "Video processing using deep learning techniques: A systematic literature review," *IEEE Access*, vol. 9, pp. 139 489–139 507, 2021, doi:10.1109/ACCESS.2021.3118541.

[51] F. L. Sánchez, I. Hupont, S. Tabik, and F. Herrera, "Revisiting crowd behaviour analysis through deep learning: Taxonomy, anomaly detection, crowd emotions, datasets, opportunities and prospects," *Information Fusion*, vol. 64, pp. 318–335, 2020, doi:10.1016/j.inffus.2020.07.008.

[52] X. Ran, H. Chen, X. Zhu, Z. Liu, and J. Chen, "Deepdecision: A mobile deep learning framework for edge video analytics," in *IEEE INFOCOM 2018-IEEE conference on computer communications*, pp. 1421–1429, doi:10.1109/INFOCOM.2018.8485905.

[53] L. Deng, G. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: An overview," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8599–8603, doi:10.1109/ICASSP.2013.6639344.

[54] A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, "Speech recognition using deep neural networks: A systematic review," *IEEE access*, vol. 7, pp. 19 143–19 165, 2019, doi:10.1109/ACCESS.2019.2896880.

[55] Z. Zhang, J. Geiger, J. Pohjalainen, A. E.-D. Mousa, W. Jin, and B. Schuller, "Deep learning for environmentally robust speech recognition: An overview of recent developments," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 9, no. 5, pp. 1–28, 2018, doi:10.1145/3178115.

[56] D. W. Otter, J. R. Medina, and J. K. Kalita, "A survey of the usages of deep learning for natural language processing," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 604–624, 2020, doi:10.1109/TNNLS.2020.2979670.

[57] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *IEEE Computational Intelligence Magazine*, vol. 13, no. 3, pp. 55–75, 2018, doi:10.1109/MCI.2018.2840738.

[58] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 160–167, doi:10.1145/1390156.1390177.

[59] Y. LeCun, Y. Bengio *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995, doi:10.5555/303568.303704.

[60] L. R. Medsker and L. Jain, "Recurrent neural networks," *Design and Applications*, vol. 5, no. 64-67, p. 2, 2001.

[61] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, and L. Bottou, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of machine learning research*, vol. 11, no. 12, 2010, doi:10.5555/1756006.1953039.

[62] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis *et al.*, "Deep learning for computer vision: A brief review," *Computational intelligence and neuroscience*, vol. 2018, 2018, doi:10.1155/2018/7068349.

[63] M. Hassaballah and A. I. Awad, *Deep learning in computer vision: principles and applications*. CRC Press, 2020.

[64] Y. Wu, D. Lian, Y. Xu, L. Wu, and E. Chen, "Graph convolutional networks with markov random field reasoning for social spammer detection," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 01, 2020, pp. 1054–1061, doi:10.1609/aaai.v34i01.5455.

[65] T. Hamaguchi, H. Oiwa, M. Shimbo, and Y. Matsumoto, "Knowledge transfer for out-of-knowledge-base entities: A graph neural network approach," *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pp. 1802–1808, 2017, doi:10.24963/ijcai.2017/250.

[66] A. Fout, J. Byrd, B. Shariat, and A. Ben-Hur, "Protein interface prediction using graph convolutional networks," *Advances in neural information processing systems*, vol. 30, 2017, doi:10.5555/3295222.3295399.

[67] U. A. Bhatti, H. Tang, G. Wu, S. Marjan, and A. Hussain, "Deep learning with graph convolutional networks: An overview and latest applications in computational intelligence," *International Journal of Intelligent Systems*, vol. 2023, pp. 1–28, 2023, doi:10.1155/2023/8342104.

[68] Y. Rong, T. Xu, J. Huang, W. Huang, H. Cheng, Y. Ma, Y. Wang, T. Derr, L. Wu, and T. Ma, "Deep graph learning: Foundations, advances and applications," in *Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2020, pp. 3555–3556, doi:10.1145/3394486.3406474.

[69] F. Xia, J. Wang, X. Kong, D. Zhang, and Z. Wang, "Ranking station importance with human mobility patterns using subway network datasets," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 7, pp. 2840–2852, 2019, doi:10.1109/TITS.2019.2920962.

[70] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61–80, 2008, doi:10.1109/TNN.2008.2005605.

[71] L. Wu, P. Cui, J. Pei, L. Zhao, and X. Guo, "Graph neural networks: foundation, frontiers and applications," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 4840–4841, doi:10.1145/3580305.3599560.

[72] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2020, doi:10.1109/TNNLS.2020.2978386.

[73] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *International Conference on Learning Representations*, 2018.

[74] D. Beck, G. Haffari, and T. Cohn, "Graph-to-sequence learning using gated graph neural networks," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 273–283, doi:10.18653/v1/P18-1026.

[75] X. Liang, X. Shen, J. Feng, L. Lin, and S. Yan, "Semantic object parsing with graph lstm," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 125–143, doi:10.1007/978-3-319-46448-0_8.

[76] K. Xu, C. Li, Y. Tian, T. Sonobe, K.-i. Kawarabayashi, and S. Jegelka, "Representation learning on graphs with jumping knowledge networks," in *Proceedings of the 35th International Conference on Machine Learning*, 2018, pp. 5453–5462.

[77] Z. Ying, J. You, C. Morris, X. Ren, W. Hamilton, and J. Leskovec, "Hierarchical graph representation learning with differentiable pooling," *Advances in Neural Information Processing Systems*, vol. 31, 2018, doi:10.5555/3327345.3327389.

[78] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," *2nd International Conference on Learning Representations*, 2014, doi:10.48550/arXiv.1312.6203.

[79] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," *Advances in neural information processing systems*, vol. 29, 2016, doi:10.5555/3157382.3157527.

[80] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011, doi:10.1016/j.acha.2010.04.005.

[81] H. Gao, Z. Wang, and S. Ji, "Large-scale learnable graph convolutional networks," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 1416–1424, doi:10.1145/3219819.3219947.

[82] J. Du, S. Zhang, G. Wu, J. M. Moura, and S. Kar, "Topology adaptive graph convolutional networks," *arXiv preprint arXiv:1710.10370*, 2017, doi:10.48550/arXiv.1710.10370.

[83] T. Yao, Y. Pan, Y. Li, and T. Mei, "Exploring visual relationship for image captioning," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 684–699, doi:10.1007/978-3-030-01264-9$_4$2.

[84] J. Johnson, A. Gupta, and L. Fei-Fei, "Image generation from scene graphs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 1219–1228, doi:10.1109/CVPR.2018.00133.

[85] H. Chen, M. Li, L. Jing, and Z. Cheng, "Lightweight long and short-range spatial-temporal graph convolutional network for skeleton-based action recognition," *IEEE Access*, vol. 9, pp. 161 374–161 382, 2021, doi:10.1109/ACCESS.2021.3131809.

[86] X. Gao, W. Hu, J. Tang, P. Pan, J. Liu, and Z. Guo, "Generalized graph convolutional networks for skeleton-based action recognition," *arXiv preprint arXiv:1811.12013*, vol. 3, 2018.

[87] C. Wang, B. Samari, and K. Siddiqi, "Local spectral graph convolution for point set feature learning," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 52–66, doi:10.1007/978-3-030-01225-0$_4$.

[88] N. Verma, E. Boyer, and J. Verbeek, "Feastnet: Feature-steered graph convolutions for 3D shape analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 2598–2606, doi:10.1109/CVPR.2018.00275.

[89] L. Yao, C. Mao, and Y. Luo, "Graph convolutional networks for text classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 7370–7377, doi:10.1609/aaai.v33i01.33017370.

[90] Y. Lin, Y. Meng, X. Sun, Q. Han, K. Kuang, J. Li, and F. Wu, "Bertgcn: Transductive text classification by combining gnn and bert," in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pp. 1456–1462, doi:10.18653/v1/2021.findings-acl.126.

[91] Y. Zhang, P. Qi, and C. D. Manning, "Graph convolution over pruned dependency trees improves relation extraction," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 2205–2215, doi:10.18653/v1/D18-1244.

[92] T. Nguyen and R. Grishman, "Graph convolutional networks with argument-aware pooling for event detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018, doi:10.5555/3504035.3504759.

[93] D. Mrowca, C. Zhuang, E. Wang, N. Haber, L. F. Fei-Fei, J. Tenenbaum, and D. L. Yamins, "Flexible neural representation for physics prediction," *Advances in Neural Information Processing Systems*, vol. 31, 2018, doi:10.5555/3327546.3327557.

[94] X. Li, X. Yan, Q. Gu, H. Zhou, D. Wu, and J. Xu, "Deepchemstable: chemical stability prediction with an attention-based graph convolution network," *Journal of chemical information and modeling*, vol. 59, no. 3, pp. 1044–1049, 2019, doi:10.1021/acs.jcim.8b00672.

[95] Y. Huang, S. Wuchty, Y. Zhou, and Z. Zhang, "SGPPI: structure-aware prediction of protein–protein interactions in rigorous conditions with graph convolutional network," *Briefings in Bioinformatics*, vol. 24, no. 2, p. bbad020, 2023, doi:10.1093/bib/bbad020.

[96] S. Sunny, P. B. Prakash, G. Gopakumar, and P. Jayaraj, "DeepBindPPI: Protein–protein binding site prediction using attention based graph convolutional network," *The Protein Journal*, pp. 1–12, 2023, doi:10.1007/s10930-023-10121-9.

[97] M. Li, H. Chen, Y. Liu, and Q. Zhao, "4s-SleepGCN: Four-stream graph convolutional networks for sleep stage classification," *IEEE Access*, vol. 11, pp. 70 621–70 634, 2023, doi:10.1109/ACCESS.2023.3294410.

[98] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018, doi:10.5555/3504035.3504947.

[99] V. Eramo, F. Lavacca, F. Valente, V. Filippetti, A. Rosato, A. Verdone, and M. Panella, "Neural graphs: an effective solution for the resource allocation in nfv sites interconnected by elastic optical networks," in *2023 23rd International Conference on Transparent Optical Networks (ICTON)*. IEEE, pp. 1–6, doi:10.1109/ICTON59386.2023.10207206.

[100] C. Yang and G. Qi, "An urban traffic knowledge graph-driven spatial-temporal graph convolutional network for traffic flow prediction," in *Proceedings of the 11th International Joint Conference on Knowledge Graphs*, 2022, pp. 110–114, doi:10.1145/3579051.3579058.

[101] X. Wu, H. Chen, R. Jin, and Q. Ni, "Spatial temporal graph convolutional network model for rumor source detection under multiple observations in social networks," in *International Wireless Internet Conference*. Springer, 2022, pp. 201–212, doi:10.1007/978-3-031-27041-3$_1$4.

[102] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018, doi:10.48550/arXiv.1803.01271.

[103] M. Niepert, M. Ahmed, and K. Kutzkov, "Learning convolutional neural networks for graphs," in *Proceedings of the 33rd International Conference on International Conference on Machine Learning*, 2016, pp. 2014–2023, doi:10.5555/3045390.3045603.

[104] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mechanical systems and signal processing*, vol. 151, p. 107398, 2021, doi:10.1016/j.ymssp.2020.107398.

[105] J. Zhang, G. Ye, Z. Tu, Y. Qin, Q. Qin, J. Zhang, and J. Liu, "A spatial attentive and temporal dilated (SATD) GCN for skeleton-based action recognition," *CAAI Transactions on Intelligence Technology*, vol. 7, no. 1, pp. 46–55, 2022, doi:10.1049/cit2.12012.

[106] B. Li, X. Li, Z. Zhang, and F. Wu, "Spatio-temporal graph routing for skeleton-based action recognition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 8561–8568, doi:10.1609/aaai.v33i01.33018561.

[107] M. Li, S. Chen, X. Chen, Y. Zhang, Y. Wang, and Q. Tian, "Actional-structural graph convolutional networks for skeleton-based action recognition," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3595–3603, doi:10.1109/CVPR.2019.00371.

[108] L. Shi, Y. Zhang, J. Cheng, and H. Lu, "Skeleton-based action recognition with directed graph neural networks," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7912–7921, doi:10.1109/CVPR.2019.00810.

[109] ——, "Two-stream adaptive graph convolutional networks for skeleton-based action recognition," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12 026–12 035, doi:10.1109/CVPR.2019.01230.

[110] C. Si, W. Chen, W. Wang, L. Wang, and T. Tan, "An attention enhanced graph convolutional LSTM network for skeleton-based action recognition," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1227–1236, doi:10.1109/CVPR.2019.00132.

[111] Y.-H. Wen, L. Gao, H. Fu, F.-L. Zhang, and S. Xia, "Graph CNNs with motif and variable temporal block for skeleton-based action recognition," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 8989–8996, doi:10.1609/aaai.v33i01.33018989.

[112] M. Wu and P. Shi, "Human pose estimation based on a spatial temporal graph convolutional network," *Applied Sciences*, vol. 13, no. 5, p. 3286, 2023, doi:10.3390/app13053286.

[113] Y. Cai, L. Ge, J. Liu, J. Cai, T.-J. Cham, J. Yuan, and N. M. Thalmann, "Exploiting spatial-temporal relationships for 3D pose estimation via graph convolutional networks," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2272–2281, doi:10.1109/ICCV.2019.00236.

[114] T. Sofianos, A. Sampieri, L. Franco, and F. Galasso, "Space-time-separable graph convolutional network for pose forecasting," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 11 209–11 218, doi:10.1109/ICCV48922.2021.01102.

[115] J. Liu, J. Rojas, Y. Li, Z. Liang, Y. Guan, N. Xi, and H. Zhu, "A graph attention spatio-temporal convolutional network for 3D human pose estimation in video," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3374–3380, doi:10.1109/ICRA48506.2021.9561605.

[116] O. Keskes and R. Noumeir, "Vision-based fall detection using ST-GCN," *IEEE Access*, vol. 9, pp. 28 224–28 236, 2021, doi:10.1109/ACCESS.2021.3058219.

[117] D. N. Tien, V. Do Hoang, and T. N. Van, "Vision-based fall detection system for the elderly using image processing and deep learning," in *Third International Conference on Computer Vision and Information Technology (CVIT 2022)*, vol. 12590. SPIE, 2023, pp. 16–26, doi:10.1117/12.2670053.

[118] P. Lu, W. Bai, D. Rueckert, and J. A. Noble, "Dynamic spatio-temporal graph convolutional networks for cardiac motion analysis," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pp. 122–125, doi:10.1109/ISBI48211.2021.9433890.

[119] M. Li, H. Chen, and Z. Cheng, "An attention-guided spatiotemporal graph convolutional network for sleep stage classification," *Life*, vol. 12, no. 5, p. 622, 2022, doi:10.3390/life12050622.

[120] Y. Zhao, X. Lin, Z. Zhang, X. Wang, X. He, and L. Yang, "STDP-based adaptive graph convolutional networks for automatic sleep staging," *Frontiers in Neuroscience*, vol. 17, p. 1158246, 2023, doi:10.3389/fnins.2023.1158246.

[121] D. K. Ghosh, A. Chakrabarty, H. Moon, and M. J. Piran, "A spatio-temporal graph convolutional network model for internet of medical things (IoMT)," *Sensors*, vol. 22, no. 21, p. 8438, 2022, doi:10.3390/s22218438.

[122] R. A. Rensink, "The dynamic representation of scenes," *Visual Cognition*, vol. 7, no. 1-3, pp. 17–42, 2000, doi:10.1080/135062800394667.

[123] M. Corbetta and G. L. Shulman, "Control of goal-directed and stimulus-driven attention in the brain," *Nature Reviews Neuroscience*, vol. 3, no. 3, pp. 201–215, 2002, doi:10.1038/nrn755.

[124] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. R. Martin, M.-M. Cheng, and S.-M. Hu, "Attention mechanisms in computer vision: A survey," *Computational Visual Media*, vol. 8, no. 3, pp. 331–368, 2022, doi:10.1007/s41095-022-0271-y.

[125] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998, doi:10.1109/34.730558.

[126] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141, doi:10.1109/CVPR.2018.00745.

[127] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European Conference on Computer Vision*. Springer, 2020, pp. 213–229, doi:10.1007/978-3-030-58452-8_13.

[128] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3141–3149, doi:10.1109/CVPR.2019.00326.

[129] J. Yang, P. Ren, D. Zhang, D. Chen, F. Wen, H. Li, and G. Hua, "Neural aggregation network for video face recognition," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4362–4371, doi:10.1109/CVPR.2017.554.

[130] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11 065–11 074, doi:10.1109/CVPR.2019.01132.

[131] Q. Guan, Y. Huang, Z. Zhong, Z. Zheng, L. Zheng, and Y. Yang, "Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification," *arXiv preprint arXiv:1801.09927*, 2018, doi:10.48550/arXiv.1801.09927.

[132] T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, and X. He, "AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1316–1324, doi:10.1109/CVPR.2018.00143.

[133] V. Mnih, N. Heess, A. Graves *et al.*, "Recurrent models of visual attention," *Advances in Neural Information Processing Systems*, vol. 27, 2014, doi:10.5555/2969033.2969073.

[134] M. Jaderberg, K. Simonyan, A. Zisserman *et al.*, "Spatial transformer networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015, doi:10.5555/2969442.2969465.

[135] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7794–7803, doi:10.1109/CVPR.2018.00813.

[136] J. Hu, L. Shen, S. Albanie, G. Sun, and A. Vedaldi, "Gather-excite: Exploiting feature context in convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 31, 2018, doi:10.5555/3327546.3327612.

[137] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi:10.1162/neco.1997.9.8.1735.

[138] R. S. Sutton, D. McAllester, S. Singh, and Y. Mansour, "Policy gradient methods for reinforcement learning with function approximation," *Advances in Neural Information Processing Systems*, vol. 12, 1999, doi:10.5555/3009657.3009806.

[139] K. Gregor, I. Danihelka, A. Graves, D. Rezende, and D. Wierstra, "Draw: A recurrent neural network for image generation," in *Proceedings of the 32nd International Conference on Machine Learning*, 2015, pp. 1462–1471, doi:10.48550/arXiv.1502.04623.

[140] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *Proceedings of the 32nd International Conference on Machine Learning*, vol. 37, 2015, pp. 2048–2057, doi:10.5555/3045118.3045336.

[141] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "CCNet: Criss-cross attention for semantic segmentation," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 603–612, doi:10.1109/ICCV.2019.00069.

[142] Y. Chen, M. Rohrbach, Z. Yan, Y. Shuicheng, J. Feng, and Y. Kalantidis, "Graph-based global reasoning networks," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 433–442, doi:10.1109/CVPR.2019.00052.

[143] H. Zhao, J. Jia, and V. Koltun, "Exploring self-attention for image recognition," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10 073–10 082, doi:10.1109/CVPR42600.2020.01009.

[144] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *International Conference on Learning Representations*, 2020, doi:10.48550/arXiv.2010.11929.

[145] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, and T.-S. Chua, "SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6298–6306, doi:10.1109/CVPR.2017.667.

[146] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015, doi:10.1109/JPROC.2015.2460697.

[147] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli, "Multimodal fusion for multimedia analysis: a survey," *Multimedia Systems*, vol. 16, pp. 345–379, 2010, doi:10.1007/s00530-010-0182-0.

[148] D. Ramachandram and G. W. Taylor, "Deep multimodal learning: A survey on recent advances and trends," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 96–108, 2017, doi:10.1109/MSP.2017.2738401.

[149] S. Poria, E. Cambria, and A. Gelbukh, "Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 2539–2544, doi:10.18653/v1/D15-1303.

[150] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006, doi:10.1126/science.1127647.

[151] H. P. Martínez and G. N. Yannakakis, "Deep multimodal fusion: Combining discrete events and continuous signals," in *Proceedings of the 16th International Conference on Multimodal Interaction*, 2014, pp. 34–41, doi:10.1145/2663204.2663236.

[152] A. Eitel, J. T. Springenberg, L. Spinello, M. Riedmiller, and W. Burgard, "Multimodal deep learning for robust RGB-D object recognition," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 681–687, doi:10.1109/IROS.2015.7353446.

[153] N. Srivastava and R. R. Salakhutdinov, "Multimodal learning with deep Boltzmann machines," *Advances in Neural Information Processing Systems*, vol. 25, 2012, doi:10.5555/2627435.2697059.

[154] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, 2021, doi:10.1016/j.neucom.2021.03.091.

[155] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1725–1732, doi:10.1109/CVPR.2014.223.

[156] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proceedings of the 28th International Conference on International Conference on Machine Learning*, pp. 689–696, doi:10.5555/3104482.3104569.

[157] J. Liu, A. Shahroudy, D. Xu, A. C. Kot, and G. Wang, "Skeleton-based action recognition using spatio-temporal LSTM network with trust gates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 12, pp. 3007–3021, 2017, doi:10.1109/TPAMI.2017.2771306.

[158] Q. Miao, Y. Li, W. Ouyang, Z. Ma, X. Xu, W. Shi, and X. Cao, "Multimodal gesture recognition based on the ResC3D network," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 3047–3055, doi:10.1109/ICCVW.2017.360.

[159] G. K. Verma and U. S. Tiwary, "Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals," *NeuroImage*, vol. 102, pp. 162–172, 2014, doi:10.1016/j.neuroimage.2013.11.007.

[160] C. Ding and D. Tao, "Robust face recognition via multimodal deep face representation," *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp. 2049–2058, 2015, doi:10.1109/TMM.2015.2477042.

[161] D. J. Loeffelbein, M. Souvatzoglou, V. Wankerl, A. Martinez-Möller, J. Dinges, M. Schwaiger, and A. J. Beer, "PET-MRI fusion in head-and-neck oncology: current status and implications for hybrid PET/MRI," *Journal of Oral and Maxillofacial Surgery*, vol. 70, no. 2, pp. 473–483, 2012, doi:10.1016/j.joms.2011.02.120.

[162] F. Mohsen, H. Ali, N. El Hajj, and Z. Shah, "Artificial intelligence-based methods for fusion of electronic health records and imaging data," *Scientific Reports*, vol. 12, no. 1, p. 17981, 2022, doi:10.1038/s41598-022-22514-4.

[163] J. Ye, J. Hai, J. Song, and Z. Wang, "Multimodal data hybrid fusion and natural language processing for clinical prediction models," *medRxiv*, pp. 2023–08, 2023, doi:10.1101/2023.08.24.23294597.

[164] K. Wulff, S. Gatti, J. G. Wettstein, and R. G. Foster, "Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease," *Nature Reviews Neuroscience*, vol. 11, no. 8, pp. 589–599, 2010, doi:10.1038/nrn2868.

[165] E. Estrada, H. Nazeran, J. Barragan, J. R. Burk, E. A. Lucas, and K. Behbehani, "EOG and EMG: Two important switches in automatic sleep stage classification," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2458–2461, doi:10.1109/IEMBS.2006.260075.

[166] A. Krakovská and K. Mezeiová, "Automatic sleep scoring: A search for an optimal combination of measures," *Artificial intelligence in medicine*, vol. 53, no. 1, pp. 25–33, 2011.

[167] L. Fiorillo, A. Puiatti, M. Papandrea, P.-L. Ratti, P. Favaro, C. Roth, P. Bargiotas, C. L. Bassetti, and F. D. Faraci, "Automated sleep scoring: A review of the latest approaches," *Sleep Medicine Reviews*, vol. 48, p. 101204, 2019, doi:10.1016/j.smrv.2019.07.007.

[168] O. Tsinalis, P. M. Matthews, and Y. Guo, "Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders," *Annals of Biomedical Engineering*, vol. 44, pp. 1587–1597, 2016, doi:10.1007/s10439-015-1444-y.

[169] P.-L. Lee, Y.-H. Huang, P.-C. Lin, Y.-A. Chiao, J.-W. Hou, H.-W. Liu, Y.-L. Huang, Y.-T. Liu, and T.-D. Chiueh, "Automatic sleep staging in patients with obstructive sleep apnea using single-channel frontal EEG," *Journal of Clinical Sleep Medicine*, vol. 15, no. 10, pp. 1411–1420, 2019, doi:10.5664/jcsm.7964.

[170] H. Dong, A. Supratak, W. Pan, C. Wu, P. M. Matthews, and Y. Guo, "Mixed neural network approach for temporal sleep stage classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 2, pp. 324–333, 2017, doi:10.1109/TNSRE.2017.2733220.

[171] K. Mikkelsen and M. De Vos, "Personalizing deep learning models for automatic sleep staging," *arXiv preprint arXiv:1801.02645*, 2018, doi:10.48550/arXiv.1801.02645.

[172] E. Alickovic and A. Subasi, "Ensemble SVM method for automatic sleep stage classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 6, pp. 1258–1265, 2018, doi:10.1109/TIM.2018.2799059.

[173] P. Memar and F. Faradji, "A novel multi-class EEG-based sleep stage classification system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 1, pp. 84–95, 2017, doi:10.1109/TNSRE.2017.2776149.

[174] A. Sikka, H. Jamalabadi, M. Krylova, S. Alizadeh, J. N. van der Meer, L. Danyeli, M. Deliano, P. Vicheva, T. Hahn, T. Koenig *et al.*, "Investigating the temporal dynamics of electroencephalogram (EEG) microstates using recurrent neural networks," *Human Brain Mapping*, vol. 41, no. 9, pp. 2334–2346, 2020, doi:10.1002/hbm.24949.

[175] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. De Vos, "SeqSleepNet: end-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 3, pp. 400–410, 2019, doi:10.1109/TNSRE.2019.2896659.

[176] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734, doi:10.3115/v1/D14-1179.

[177] H. Seo, S. Back, S. Lee, D. Park, T. Kim, and K. Lee, "Intra-and inter-epoch temporal context network (IITNet) using sub-epoch features for automatic sleep scoring on raw single-channel EEG," *Biomedical Signal Processing and Control*, vol. 61, p. 102037, 2020, doi:10.1016/j.bspc.2020.102037.

[178] O. Tsinalis, P. M. Matthews, Y. Guo, and S. Zafeiriou, "Automatic sleep stage scoring with single-channel EEG using convolutional neural networks," *arXiv preprint arXiv:1610.01683*, 2016, doi:10.48550/arXiv.1610.01683.

[179] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. De Vos, "Joint classification and prediction CNN framework for automatic sleep stage classification," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 5, pp. 1285–1296, 2018, doi:10.1109/TBME.2018.2872652.

[180] S. Mousavi, F. Afghah, and U. R. Acharya, "SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach," *PloS One*, vol. 14, no. 5, p. e0216456, 2019, doi:10.1371/journal.pone.0216456.

[181] S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, and A. Gramfort, "A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 4, pp. 758–769, 2018, doi:10.1109/TNSRE.2018.2813138.

[182] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 1998–2008, 2017, doi:10.1109/TNSRE.2017.2721116.

[183] A. Supratak and Y. Guo, "TinySleepNet: An efficient deep learning model for sleep stage scoring based on raw single-channel EEG," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 641–644, doi:10.1109/EMBC44109.2020.9176741.

[184] S. Zhang, D. Chen, Y. Tang, and L. Zhang, "Children ASD evaluation through joint analysis of EEG and eye-tracking recordings with graph convolution network," *Frontiers in Human Neuroscience*, vol. 15, p. 651349, 2021, doi:10.3389/fnhum.2021.651349.

[185] Z. Jia, Y. Lin, J. Wang, R. Zhou, X. Ning, Y. He, and Y. Zhao, "GraphSleepNet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification." in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)*, vol. 2021, 2020, pp. 1324–1330, doi:10.24963/ijcai.2020/184.

[186] Z. Jia, Y. Lin, J. Wang, X. Ning, Y. He, R. Zhou, Y. Zhou, and H. L. Li-wei, "Multi-view spatial-temporal graph convolutional networks with domain generalization for sleep stage classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 1977–1986, 2021, doi:10.1109/TNSRE.2021.3110665.

[187] A. Hyvärinen, "New approximations of differential entropy for independent component analysis and projection pursuit," *Advances in Neural Information Processing Systems*, vol. 10, 1997, doi:10.5555/3008904.3008943.

[188] A. Sagheer and M. Kotb, "Time series forecasting of petroleum production using deep LSTM recurrent networks," *Neurocomputing*, vol. 323, pp. 203–213, 2019, doi:10.1016/j.neucom.2018.09.082.

[189] S. Ebrahimi Kahou, V. Michalski, K. Konda, R. Memisevic, and C. Pal, "Recurrent neural networks for emotion recognition in video," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pp. 467–474, doi:10.1145/2818346.2830596.

[190] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *Scientific Reports*, vol. 8, no. 1, p. 6085, 2018, doi:10.1038/s41598-018-24271-9.

[191] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018, doi:10.48550/arXiv.1803.01271.

[192] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-Resnet and the impact of residual connections on learning," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, doi:10.1609/aaai.v31i1.11231.

[193] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *International Conference on Learning Representations*, 2016, doi:10.48550/arXiv.1511.07122.

[194] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, doi:10.1109/CVPR.2016.90.

[195] F. Baradel, C. Wolf, and J. Mille, "Human action recognition: Pose-based attention draws focus to hands," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017, pp. 604–613, doi:10.1109/ICCVW.2017.77.

[196] Y. Liu, Z. Zhang, X. Liu, W. Lei, and X. Xia, "Deep learning based mineral image classification combined with visual attention mechanism," *IEEE Access*, vol. 9, pp. 98 091–98 109, 2021, doi:10.1109/ACCESS.2021.3095368.

[197] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 13 708–13 717, doi:10.1109/CVPR46437.2021.01350.

[198] N. Oliver, G. Smith, C. Thakkar, and A. C. Surendran, "Swish: semantic analysis of window titles and switching history," in *Proceedings of the 11th international conference on Intelligent user interfaces*, 2006, pp. 194–201, doi:10.1109/ACCESS.2021.3095368.

[199] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. Kamphuisen, and J. J. Oberye, "Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 9, pp. 1185–1194, 2000, doi:10.1109/10.867928.

[200] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000, doi:10.1161/01.cir.101.23.e215.

[201] S. Khalighi, T. Sousa, J. M. Santos, and U. Nunes, "ISRUC-Sleep: A comprehensive public dataset for sleep researchers," *Computer methods and programs in biomedicine*, vol. 124, pp. 180–192, 2016, doi:10.1016/j.cmpb.2015.10.013.

[202] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006, doi:10.1016/j.patrec.2005.10.010.

[203] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960, doi:10.1177/001316446002000104.

[204] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009, doi:10.1016/j.ipm.2009.03.002.

[205] X. Chen, J. He, X. Wu, W. Yan, and W. Wei, "Sleep staging by bidirectional long short-term memory convolution neural network," *Future Generation Computer Systems*, vol. 109, pp. 188–196, 2020, doi:10.1016/j.future.2020.03.019.

[206] H. Ghimatgar, K. Kazemi, M. S. Helfroush, and A. Aarabi, "An automatic single-channel EEG-based sleep stage scoring method based on hidden markov model," *Journal of Neuroscience Methods*, vol. 324, p. 108320, 2019, doi:10.1016/j.jneumeth.2019.108320.

[207] H. Shen, F. Ran, M. Xu, A. Guez, A. Li, and A. Guo, "An automatic sleep stage classification algorithm using improved model based essence features," *Sensors*, vol. 20, no. 17, p. 4677, 2020, doi:10.3390/s20174677.

[208] M. M. Ohayon, "Epidemiology of insomnia: what we know and what we still need to learn," *Sleep Medicine Reviews*, vol. 6, no. 2, pp. 97–111, 2002, doi:10.1053/smrv.2002.0186.

[209] S. J. Schreiner, L. L. Imbach, P. O. Valko, A. Maric, R. Maqkaj, E. Werth, C. R. Baumann, and H. Baumann-Vogel, "Reduced regional NREM sleep slow-wave activity is associated with cognitive impairment in Parkinson disease," *Frontiers in Neurology*, vol. 12, p. 618101, 2021, doi:10.3389/fneur.2021.618101.

[210] L. Tafaro, P. Cicconetti, A. Baratta, N. Brukner, E. Ettorre, V. Marigliano, and M. Cacciafesta, "Sleep quality of centenarians: cognitive and survival implications," *Archives of Gerontology and Geriatrics*, vol. 44, pp. 385–389, 2007, doi:10.1016/j.archger.2007.01.054.

[211] S. A. Joosten, S. A. Landry, A.-M. Wong, D. L. Mann, P. I. Terrill, S. A. Sands, A. Turton, C. Beatty, L. Thomson, G. S. Hamilton *et al.*, "Assessing the physiologic endotypes responsible for REM- and NREM-based OSA," *Chest*, vol. 159, no. 5, pp. 1998–2007, 2021, doi:10.1016/j.chest.2020.10.080.

[212] R. Ren, N. Covassin, Y. Zhang, F. Lei, L. Yang, J. Zhou, L. Tan, T. Li, Y. Li, J. Shi *et al.*, "Interaction between slow wave sleep and obstructive sleep apnea in prevalent hypertension," *Hypertension*, vol. 75, no. 2, pp. 516–523, 2020, doi:10.1161/HYPERTENSIONAHA.119.13720 .

[213] D. S. Modha, R. Ananthanarayanan, S. K. Esser, A. Ndirango, A. J. Sherbondy, and R. Singh, "Cognitive computing," *Communications of the ACM*, vol. 54, no. 8, pp. 62–71, 2011.

[214] J. M. Siegel, "Clues to the functions of mammalian sleep," *Nature*, vol. 437, no. 7063, pp. 1264–1271, 2005, doi:10.1038/nature04285.

[215] P. H. Finan, P. J. Quartana, B. Remeniuk, E. L. Garland, J. L. Rhudy, M. Hand, M. R. Irwin, and M. T. Smith, "Partial sleep deprivation attenuates the positive affective system: effects across multiple measurement modalities," *Sleep*, vol. 40, no. 1, p. zsw017, 2017, doi:10.1093/sleep/zsw017.

[216] T. Young, P. E. Peppard, and D. J. Gottlieb, "Epidemiology of obstructive sleep apnea: a population health perspective," *American Journal of Respiratory and Critical Care Medicine*, vol. 165, no. 9, pp. 1217–1239, 2002, doi:10.1164/rccm.2109080.

[217] H. Danker-Hopfe, D. Kunz, G. Gruber, G. Klösch, J. L. Lorenzo, S.-L. Himanen, B. Kemp, T. Penzel, J. Röschke, H. Dorn *et al.*, "Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders," *Journal of Sleep Research*, vol. 13, no. 1, pp. 63–69, 2004, doi:10.1046/j.1365-2869.2003.00375.x.

[218] G. Zhu, Y. Li, and P. Wen, "Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 6, pp. 1813–1821, 2014, doi:10.1109/JBHI.2014.2303991.

[219] H. G. Jo, J. Y. Park, C. K. Lee, S. K. An, and S. K. Yoo, "Genetic fuzzy classifier for sleep stage identification," *Computers in Biology and Medicine*, vol. 40, no. 7, pp. 629–634, 2010, doi:10.1016/j.compbiomed.2010.04.007.

[220] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," *arXiv preprint arXiv:1409.2329*, 2014, doi:10.48550/arXiv.1409.2329.

[221] E. Bresch, U. Großekathöfer, and G. Garcia-Molina, "Recurrent deep neural networks for real-time sleep stage classification from single channel EEG," *Frontiers in Computational Neuroscience*, vol. 12, p. 85, 2018, doi:10.3389/fncom.2018.00085.

[222] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. De Vos, "Automatic sleep stage classification using single-channel EEG: Learning sequential features with attention-based recurrent neural networks," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1452–1455, doi:10.1109/EMBC.2018.8512480.

[223] A. Sors, S. Bonnet, S. Mirek, L. Vercueil, and J.-F. Payen, "A convolutional neural network for sleep stage scoring from raw single-channel EEG," *Biomedical Signal Processing and Control*, vol. 42, pp. 107–114, 2018, doi:10.1016/j.bspc.2017.12.001.

[224] Y. Fang, Y. Xia, P. Chen, J. Zhang, and Y. Zhang, "A dual-stream deep neural network integrated with adaptive boosting for sleep staging," *Biomedical Signal Processing and Control*, vol. 79, p. 104150, 2023, doi:10.1016/j.bspc.2022.104150.

[225] W. Xia, T. Wang, Q. Gao, M. Yang, and X. Gao, "Graph embedding contrastive multi-modal representation learning for clustering," *IEEE Transactions on Image Processing*, vol. 32, pp. 1170–1183, 2023, doi:10.1109/TIP.2023.3240863.

[226] Z. Chen, L. Fu, J. Yao, W. Guo, C. Plant, and S. Wang, "Learnable graph convolutional network and feature fusion for multi-view learning," *Information Fusion*, vol. 95, pp. 109–119, 2023, doi:10.1016/j.inffus.2023.02.013.

[227] J. D. Geyer, S. Talathi, and P. R. Carney, "Introduction to sleep and polysomnography," *Clinical Sleep Disorders. Philadelphia: Lippincott Williams & Wilkins*, pp. 265–266, 2009.

[228] M. H. Silber, S. Ancoli-Israel, M. H. Bonnet, S. Chokroverty, M. M. Grigg-Damberger, M. Hirshkowitz, S. Kapen, S. A. Keenan, M. H. Kryger, T. Penzel *et al.*, "The visual scoring of sleep in adults," *Journal of Clinical Sleep Medicine*, vol. 3, no. 02, pp. 121–131, 2007, doi:10.5664/jcsm.26814.

[229] S. Paisarnsrisomsuk, M. Sokolovsky, F. Guerrero, C. Ruiz, and S. A. Alvarez, "Deep Sleep: Convolutional neural networks for predictive modeling of human sleep time-signals," *Proc. KDD Deep Learn. Day*, pp. 1–10, 2018.

[230] F. Andreotti, H. Phan, N. Cooray, C. Lo, M. T. Hu, and M. De Vos, "Multichannel sleep stage classification and transfer learning using convolutional neural networks," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 171–174, doi:10.1109/EMBC.2018.8512214.

[231] Z. Jia, Y. Lin, J. Wang, X. Wang, P. Xie, and Y. Zhang, "SalientSleepNet: Multimodal salient wave detection network for sleep staging," *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pp. 2614–2620, 2021, doi:10.24963/ijcai.2021/360.

[232] Z. Jia, X. Cai, G. Zheng, J. Wang, and Y. Lin, "SleepPrintNet: A multivariate multimodal neural network based on physiological time-series for automatic sleep staging," *IEEE Transactions on Artificial Intelligence*, vol. 1, no. 3, pp. 248–257, 2020, doi:10.1109/TAI.2021.3060350.

[233] Z. Jia, X. Cai, and Z. Jiao, "Multi-modal physiological signals based squeeze-and-excitation network with domain adversarial learning for sleep staging," *IEEE Sensors Journal*, vol. 22, no. 4, pp. 3464–3471, 2022, doi:10.1109/JSEN.2022.3140383.

[234] Z. Yubo, L. Yingying, Z. Bing, Z. Lin, and L. Lei, "MMASleepNet: A multimodal attention network based on electrophysiological signals for automatic sleep staging," *Frontiers in Neuroscience*, vol. 16, p. 973761, 2022, doi:10.3389/fnins.2022.973761.

[235] H. Zhu, W. Zhou, C. Fu, Y. Wu, N. Shen, F. Shu, H. Yu, C. Chen, and W. Chen, "MasksleepNet: A cross-modality adaptation neural network for heterogeneous signals processing in sleep staging," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 5, pp. 2353–2364, 2023, doi:10.1109/JBHI.2023.3253728.

[236] J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, T. Darrell, and K. Saenko, "Long-term recurrent convolutional networks for visual recognition and description," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2625–2634, doi:10.1109/CVPR.2015.7298878.

[237] Z. Chen, Z. Wu, Z. Lin, S. Wang, C. Plant, and W. Guo, "AGNN: Alternating graph-regularized neural networks to alleviate over-smoothing," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13, 2023, doi:10.1109/TNNLS.2023.3271623.

[238] P. Zhang, C. Lan, W. Zeng, J. Xing, J. Xue, and N. Zheng, "Semantics-guided neural networks for efficient skeleton-based human action recognition," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1109–1118, doi:10.1109/CVPR42600.2020.00119.

[239] J. D. M.-W. C. Kenton and L. K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, vol. 1, 2019, p. 2, doi:10.18653/v1/N19-1423.

[240] W. Peng, X. Hong, H. Chen, and G. Zhao, "Learning graph convolutional network for skeleton-based human action recognition by neural searching," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 03, 2020, pp. 2669–2676, doi:10.1609/aaai.v34i03.5652.

[241] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li, "Bag of tricks for image classification with convolutional neural networks," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 558–567, doi:10.1109/CVPR.2019.00065.

[242] J. Fan, S. Upadhye, and A. Worster, "Understanding receiver operating characteristic (ROC) curves," *Canadian Journal of Emergency Medicine*, vol. 8, no. 1, pp. 19–20, 2006, doi:10.1017/s1481803500013336.

[243] M. Perslev, M. Jensen, S. Darkner, P. J. Jennum, and C. Igel, "U-time: A fully convolutional network for time series segmentation applied to sleep staging," *Advances in Neural Information Processing Systems*, vol. 32, 2019, doi:10.5555/3454287.3454684.

[244] E. Eldele, Z. Chen, C. Liu, M. Wu, C.-K. Kwoh, X. Li, and C. Guan, "An attention-based deep learning approach for sleep stage classification with single-channel EEG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 809–818, 2021, doi:10.1109/TNSRE.2021.3076234.

[245] Y. Sun, B. Wang, J. Jin, and X. Wang, "Deep convolutional network method for automatic sleep stage classification based on neurophysiological signals," in *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1–5, doi:10.1109/CISP-BMEI.2018.8633058.

[246] X. Lv, J. Li, Q. Xu *et al.*, "A multilevel temporal context network for sleep stage classification," *Computational Intelligence and Neuroscience*, vol. 2022, 2022, doi:10.1155/2022/6104736.

[247] E. M. Wickwire, J. Geiger-Brown, S. M. Scharf, and C. L. Drake, "Shift work and shift work sleep disorder: clinical and organizational perspectives," *Chest*, vol. 151, no. 5, pp. 1156–1172, 2017, doi:10.1016/j.chest.2016.12.007.

[248] D. J. Buysse, "Sleep health: can we define it? does it matter?" *Sleep*, vol. 37, no. 1, pp. 9–17, 2014, doi:10.5665/sleep.3298.